

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:52:32 ; Search time 47 Seconds
(without alignments)
2559.888 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLSILLINTEPL.....FLSDTPQMGRSFTGGVNVKPF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4036	100.0	758	21	AA18719
2	3845	95.3	764	23	AA18719
3	3836	95.0	758	23	AAU73011
4	3819	94.6	758	21	AA18719
5	3816	94.5	764	21	AA18719
6	3814	94.5	758	21	AA18719
7	3751	92.9	758	21	AA18719
8	3714	92.0	735	21	AA18719
9	3644	90.3	728	24	AA18719

10	1808	44.8	339	21	AA18719	Neisseria meningit
11	1791	44.4	339	21	AA18719	Neisseria meningit
12	1763	43.7	339	21	AA18719	Neisseria meningit
13	1580.5	39.2	814	23	AA18719	M catarrhalis MCAI
14	1580.5	39.2	818	22	AA18719	Moraxella catarrha
15	1202	29.8	913	22	AA18719	Moraxella catarrha
16	1202	29.8	913	23	AA18719	M catarrhalis MCAI
17	763	18.9	174	21	AA18719	Neisseria meningit
18	216	5.4	713	17	AA18719	Neisseria meningit
19	216	5.4	713	22	AA18719	Neisseria meningit
20	216	5.4	713	24	AA18719	N. gonorrhoeae or N
21	214.5	5.3	720	23	AA18719	N. gonorrhoeae am
22	214	5.3	719	17	AA18719	NTHI HxuC protein
23	212.5	5.3	703	21	AA18719	Neisseria meningit
24	212.5	5.3	703	22	AA18719	Neisseria meningit
25	212.5	5.3	714	23	AA18719	Neisseria meningit
26	208.5	5.2	703	21	AA18719	Neisseria meningit
27	208.5	5.2	702	22	AA18719	Neisseria meningit
28	206	5.1	702	22	AA18719	Neisseria meningit
29	205.5	5.1	703	22	AA18719	Neisseria meningit
30	205	5.1	702	22	AA18719	Neisseria meningit
31	203	5.0	702	22	AA18719	Neisseria meningit
32	203	5.0	702	22	AA18719	Neisseria meningit
33	203	5.0	725	17	AA18719	Neisseria meningit
34	202	5.0	713	17	AA18719	Hib HxuC protein
35	202	5.0	713	22	AA18719	Neisseria gonorrh
36	202	5.0	713	23	AA18719	N. gonorrhoeae or N
37	200.5	5.0	703	22	AA18719	Neisseria gonorrh
38	200	5.0	725	20	AA18719	Neisseria meningit
39	198.5	4.9	687	22	AA18719	Pseudomonas stutze
40	198.5	4.9	708	20	AA18719	Pseudomonas ging
41	198.5	4.9	772	20	AA18719	Pseudomonas ging
42	194	4.8	757	20	AA18719	Pseudomonas ging
43	194	4.8	763	20	AA18719	Pseudomonas ging
44	189	4.7	715	17	AA18719	NTHI HxuC protein
45	189	4.7	725	20	AA18719	Neisseria meningit

ALIGNMENTS

RESULT 1

- AA18719
- ID AAB18719 standard; Protein; 758 AA.
- XX
- AC AAB18719;
- XX
- DT 22-JAN-2001 (first entry)
- XX
- DE A Neisseria meningitidis BASB082 polypeptide.
- XX
- XX
- KW BASB082; BASB083; BASB091; BASB092; infection; vaccine; gene therapy; upper respiratory tract infection; bacteremia; meningitis; invasive bacterial disease.
- KW
- XX
- OS Neisseria meningitidis.
- XX
- XX
- PN WO20005327-A2.
- XX
- PD 21-SEP-2000.
- XX
- XX
- PF 07-MAR-2000; 2000WO-EP01955.
- XX
- PR 12-MAR-1999; 99GB-0005815.
- PR 21-APR-1999; 99GB-0009094.
- PR 23-APR-1999; 99GB-0009503.
- PR 28-APR-1999; 99GB-0009787.
- PR 07-MAY-1999; 99GB-0010710.
- XX
- XX (SMIX) SMITHKLINE BEECHAM BIOLOGICALS.
- PA
- XX Defrenne C, Delmelle C, Ruelle J;
- XX

DR WPI; 2000-602119/57.
 DR N-PSDB; AAA75744.
 XX Novel polypeptides designated BASB 082, 083, 091, 092, and 101 derived
 PT from meningococcus bacterium useful for producing vaccines against
 PT infections and in diagnostic assays -
 XX Claim 3; Page 99-101; 108pp; English.
 PS
 XX The present sequence represents a BASB082 polypeptide. The specification
 CC describes BASB082, BASB091, BASB092, and BASB101 polypeptides.
 CC The polynucleotides and polypeptides are useful as diagnostic reagents.
 CC and for diagnosing N. meningitidis infection. The polynucleotides may
 CC be used as hybridisation probe for RNA, cDNA and genomic DNA to isolate
 CC full-length cDNAs and genomic clones encoding BASB082, BASB083, BASB091,
 CC BASB092 or BASB101 polypeptides and to isolate cDNA and genomic clones
 CC of other genes that have a high identity particularly high sequence
 CC identity to BASB082, BASB083, BASB091, BASB092 or BASB101 genes. The
 CC vaccine compositions are useful for inducing an immunological response
 CC in humans. The polynucleotides encoding BASB082, BASB083, BASB091,
 CC BASB092 or BASB101 polypeptides are useful in gene therapy to induce
 CC an immunological response. The polypeptides are useful for treating
 CC upper respiratory tract infection, invasive bacterial diseases, such as
 CC bacteremia and meningitis.
 XX
 SQ Sequence 758 AA;
 Query Match 100.0%; Score 4036; DB 21; Length 758;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAQTLKPIVLSILLINTPLLAQAHETEQSVGLTETVWVKSRPRATSGLLHTSTASDKI 60
 Db 1 MAQTLKPIVLSILLINTPLLAQAHETEQSVGLTETVWVKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 120
 Db 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 120
 QY 121 SPDHAIWVDLTALSQQVEILRGPTVLLYSSGNVAGLVDVADGKIPEKMPENGVSGLGLRL 180
 Db 121 SPDHAIWVDLTALSQQVEILRGPTVLLYSSGNVAGLVDVADGKIPEKMPENGVSGLGLRL 180
 QY 181 SSGNLEKLTSGGINIGLGNFVLTGELYKRSQGYAVPRYNLKRLPDSPRPFANGQHRA 240
 Db 181 SSGNLEKLTSGGINIGLGNFVLTGELYKRSQGYAVPRYNLKRLPDSPRPFANGQHRA 240
 QY 241 VLGRKKEFYERTYSDDRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTDEVD 300
 Db 241 VLGRKKEFYERTYSDDRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTDEVD 300
 QY 301 DYDNPGLSCGFHDDDDAHAAHNGKPIDLNRKYEYELRAEWKQFPFGFEALRVHLNRNDY 360
 Db 301 DYDNPGLSCGFHDDDDAHAAHNGKPIDLNRKYEYELRAEWKQFPFGFEALRVHLNRNDY 360
 QY 361 HHDEKAGDAVENFNNQTONARIELRHQPIGRKAGSGVQYLGOKSALSATSVAVKQPM 420
 Db 361 HHDEKAGDAVENFNNQTONARIELRHQPIGRKAGSGVQYLGOKSALSATSVAVKQPM 420
 QY 421 LLDNKVQHYFFGVEQANWDFNFTLEGVVRVEKQASIRYDKALIDRNYKQPLPDILGAH 480
 Db 421 LLDNKVQHYFFGVEQANWDFNFTLEGVVRVEKQASIRYDKALIDRNYKQPLPDILGAH 480
 QY 481 ROTARSFALSGNWFYTPQHKLSLTASHOERLPSTQELIYAHGKHVATNTFEVGNKHLNKR 540
 Db 481 ROTARSFALSGNWFYTPQHKLSLTASHOERLPSTQELIYAHGKHVATNTFEVGNKHLNKR 540
 QY 541 SNNIELALGVEGDRWQYNLALYRNFNGYIYAQTLNDGRGPKSTEDDSEMKLVRYNOSGA 600
 Db 541 SNNIELALGVEGDRWQYNLALYRNFNGYIYAQTLNDGRGPKSTEDDSEMKLVRYNOSGA 600
 QY 601 DFYGAEGEYVFKPTPRYRIGVSGDYVRGRKLNKLPFLGREDAYGNRPFFIAQDDQNAVRVP 660
 Db 601 DFYGAEGEYVFKPTPRYRIGVSGDYVRGRKLNKLPFLGREDAYGNRPFFIAQDDQNAVRVP 660

Db 601 DFYGAEGEYVFKPTPRYRIGVSGDYVRGRKLNKLPFLGREDAYGNRPFFIAQDDQNAVRVP 660
 QY 661 AARLGPHLKASLTDRIDANLDYRVFAQNKLARYETPTPGHMLNLGANYRRNTRYGEWN 720
 Db 661 AARLGPHLKASLTDRIDANLDYRVFAQNKLARYETPTPGHMLNLGANYRRNTRYGEWN 720
 QY 721 WYKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 Db 721 WYKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 RESULT 2
 ABB78068
 ID ABB78068 standard; Protein; 764 AA.
 XX ABB78068;
 XX 05-NOV-2002 (first entry)
 XX Amino acid sequence of p88 polypeptide.
 XX p177; p88; p64; p55; p46; vaccine; gonorrhea.
 XX Neisseria gonorrhoeae.
 XX WO200260936-A2.
 XX 08-AUG-2002.
 XX 31-JAN-2002; 2002WO-USO2881.
 XX 31-JAN-2001; 2001US-266070P.
 XX 06-AUG-2001; 2001US-310356P.
 XX 23-OCT-2001; 2001US-344452P.
 XX (TOWA) UNIV IOWA RES FOUND.
 XX (REGC) UNIV CALIFORNIA.
 XX (APIC) APICELLA M A.
 XX (EDWA) EDWARDS J L.
 XX (GIBS) GIBSON B W.
 XX (SCHE) SCHEFFLER K.
 XX (BROW) BROWN E.
 XX Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
 WPI; 2002-619227/66.
 N-PSDB; ABQ78299.
 New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
 gonorrhoeae, useful for preventing, or protecting a female patient
 against, N. gonorrhoeae colonization or infection -
 Claim 7; Page 115-117; 130pp; English.
 The present sequence represents a p88 polypeptide. The specification
 describes p177, p88, p64, p55 and p46 polypeptides from Neisseria
 gonorrhoeae. The polypeptides are useful as vaccines, for preventing,
 or protecting a female patient against, N. gonorrhoeae colonization or
 infection. Such immunisation can prevent gonorrhea in women.
 SQ Sequence 764 AA;
 Query Match 95.3%; Score 3845; DB 23; Length 764;
 Best Local Similarity 95.9%; Pred. No. 0;
 Matches 730; Conservative 7; Mismatches 18; Indels 6; Gaps 3;
 QY 1 MAQTLKPIVLSILLINTPLLAQAHETEQSVGLTETVWVKSRPRATSGLLHTSTASDKI 60
 Db 7 MAQTLKPIVLSILLINTPLLSQAHCETQSVGLTETVSVGKSRPRATSGLLHTSTASDKI 66
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 120
 Db 67 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 126

Db 658 RVPAAALGFHKLKASLTDRIDANLDYRVFAQNKILARYETRTPGHHMLNLGANYRNRTRYG 717
 QY 718 EWNWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 Db 718 EWNWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758

RESULT 4
 AAY74566
 ID AAY74566 standard; Protein; 758 AA.
 XX AC AAY74566;
 XX DE 21-MAR-2000 (first entry)
 XX DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:606.
 XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX OS Neisseria meningitidis.
 XX PN WO9957280-A2.
 XX PD 11-NOV-1999.
 XX PF 30-APR-1999; 99WU-US09346.
 XX PR 01-MAY-1998; 98US-0083758.
 XX PR 31-JUL-1998; 98US-0094869.
 XX PR 02-SEP-1998; 98US-0098994.
 XX PR 02-SEP-1998; 98US-0099062.
 XX PR 09-OCT-1998; 98US-0103749.
 XX PR 09-OCT-1998; 98US-0103794.
 XX PR 09-OCT-1998; 98US-0103796.
 XX PR 25-FEB-1999; 99US-0121528.
 XX PA (CHIR) CHIRON CORP.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 XX PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 XX PI Tettelin H, Venter JC;
 XX DR WPI; 2000-062150/05.
 XX DR N-PSDE; AAZ53328.
 XX PT Novel Neisserial polypeptides predicted to be useful antigens for
 XX PT vaccines and diagnostics -
 XX PS Claim 2; Page 424-425; 1453pp; English.
 XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 XX CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 XX CC PCR primers used in the exemplification of the present invention. The
 XX CC polypeptides, the polynucleotides, antibodies and compositions of
 XX CC the invention can be used as vaccines, as diagnostic reagents, and as
 XX CC immunogenic compositions. The polypeptides can be used in the
 XX CC manufacture of medicaments for treating or preventing infection due to
 XX CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 XX CC presence of Neisseria bacteria, or to raise antibodies. They may also
 XX CC be used to screen for agonists or antagonists, which may themselves
 XX CC have use as antibacterial agents. The polynucleotides of the invention
 XX CC may also be used in gene therapy protocols.

Query Match 94.6%; Score 3819; DB 21; Length 758;
 Best Local Similarity 95.5%; Pred. No. 0;
 Matches 72; Conservative 7; Mismatches 21; Indels 6; Gaps 3;

QY 1 MAQTTLKPIVLISILLINTPLLAQAHETEQSGLTETVTWVSKSRPRATSGLLHTSTASDKI 60
 Db 1 MAQTTLKPIVLISILLINTPLLSQAHGTEQSGLTETVSVGKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120
 Db 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAFVIRGQTGRRIKVLNHHGETGDMADF 120
 QY 121 SPDHAIWVDVLTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMPENCVSGELGURL 180
 Db 121 SPDHAIWVDVLTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMPENCVSGELGURL 180
 QY 181 SSGNLEKLTSGGINIGLGNFVLTGTEGLYRKSGDYAVPRYRNLRKLPDS--HADSTQTS 240
 Db 181 SSGNLEKLTSGGINIGLGNFVLTGTEGLYRKSGDYAVPRYRNLRKLPDS--HADSTQTS 240
 QY 241 V-LGM--RKRFPYRTYSDRDQVGLPAHSHEYDDCHADIWOKSLINKRYLQLYPHLLTE 297
 Db 238 IGLSWVGKGFIGAAYSDRDQVGLPAHSHEYDDCHADIWOKSLINKRYLQLYPHLLTE 297
 QY 298 EDVDYDNPGLSCGPHDDDDAHAAHNGKPMIDLRNKRYELRAEWKQPPGFEALRVHLNR 357
 Db 298 EDIDYDNPGLSCGPHDDDDAHAAHNGKPMIDLRNKRYELRAEWKQPPGFEALRVHLNR 357
 QY 358 NDYHHEKAGDAVENFFNNOTQNAIRIELRHQPIGRKLGSGWGVQYLGQKSSALSATSEAVK 417
 Db 358 NDYRHDEKAGDAVENFFNNOTQNAIRIELRHQPIGRKLGSGWGVQYLGQKSSALSATSEAVK 417
 QY 418 QPMLLDNKNVQHYSPFGEQANWNTLEGGRVREKOKASTRYDKALIDRENNYKQPIPLD 477
 Db 418 QPMLLDNKNVQHYSPFGEQANWNTLEGGRVREKOKASTRYDKALIDRENNYKQPIPLD 477
 QY 478 GAHQRTARSFALSGNWFYFTQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 Db 478 GAHQRTARSFALSGNWFYFTQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 QY 538 KERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLNDGRGPKSIEDDSEMKLVRYNQ 597
 Db 538 KERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLNDGRGPKSIEDDSEMKLVRYNQ 597
 QY 598 SGADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAQQDNAP 657
 Db 598 SGADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAQQDNAP 657
 QY 658 RVPAAALGFHKLKASLTDRIDANLDYRVFAQNKILARYETRTPGHHMLNLGANYRNRTRYG 717
 Db 658 RVPAAALGFHKLKASLTDRIDANLDYRVFAQNKILARYETRTPGHHMLNLGANYRNRTRYG 717
 QY 718 EWNWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 Db 718 EWNWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758

RESULT 5
 AAY74557
 ID AAY74557 standard; Protein; 764 AA.
 XX AC AAY74557;
 XX DT 21-MAR-2000 (first entry)
 XX DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:588.
 XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX OS Neisseria meningitidis.
 XX PN WO9957280-A2.
 XX PD 11-NOV-1999.

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA25437 to AA25476 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 758 AA;
 SQ Query Match 94.5%; Score 3814; DB 21; Length 758;
 Best Local Similarity 95.3%; Pred. No. 0;
 Matches 725; Conservative 10; Mismatches 20; Indels 6; Gaps 3;

QY 1 MAQTTLKPIVLSILLINTPLLAQAHEHQSVGLTVTVVVKSRPRATSGLLHTSTASDKI 60
 DB 1 MAQTTLKPIVLSILLINTPLLAQAHEHQSVGLTVTVVVKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGASAPVIRGQTRRIKVLNHHGETGDMADF 120
 DB 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGASAPVIRGQTRRIKVLNHHGETGDMADF 120
 QY 121 SPDHAIWDTALSQVEILRGVPTLLYSSGNVAGLVADGKIPEKMPGVNGVSGELGLRL 180
 DB 121 SPDHAIWDTALSQVEILRGVPTLLYSSGNVAGLVADGKIPEKMPGVNGVSGELGLRL 180
 QY 181 SSGNLEKTSGGINTGLKGNVHTEGLYKSGDVAVPRYLNKRLPDSRPFANGQHRA 240
 DB 181 SSGNLEKTSGGINTGLKGNVHTEGLYKSGDVAVPRYLNKRLPDS--HADSQTGS 237
 QY 241 V-LGW--RRFRYRTYSRDRDQYGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE 297
 DB 238 IGLSVWGEKGFIVAYSDRDRDQYGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE 297
 QY 298 EVDVDNPGSCGFHDDDAHAHANGKFWIDLNKRYELRAEWKQPPFGFALRVHLNR 357
 DB 298 EDIDVDNPGSCGFHDDDAHAHANGKFWIDLNKRYELRAEWKQPPFGFALRVHLNR 357
 QY 358 NDYHDEKAGDAVENFNNQTONARIELRHQPIGLKSGWQYVGLGQSSALSATSEAVK 417
 DB 358 NDYHDEKAGDAVENFNNQTONARIELRHQPIGLKSGWQYVGLGQSSALSATSEAVK 417
 QY 418 QPMLLDNKVQHSYFFGVEQANMDNFTLEGGRVVEKQASIRYDKALIDRENYKQPLPDL 477
 DB 418 QPMLLDNKVQHSYFFGVEQANMDNFTLEGGRVVEKQASIRYDKALIDRENYKQPLPDL 477
 QY 478 GAHQRTARSAFALSGNWFYTPQHKLSLTASHOBLPSTQELYAHGKHVATNTEVGNKHLN 537
 DB 478 GAHQRTARSAFALSGNWFYTPQHKLSLTASHOBLPSTQELYAHGKHVATNTEVGNKHLN 537
 QY 538 KERSNNIELALGVEGRWQYNLALYNNRFGNYIAQTLNDGRPKSIEDDSMKLVRYNQ 597
 DB 538 KERSNNIELALGVEGRWQYNLALYNNRFGNYIAQTLNDGRPKSIEDDSMKLVRYNQ 597
 QY 598 SGADFYGABGEIYFKPTPRYRGVSGDYVRGLKNLPSLPGREDAYGNRPFFIAQDDQNA 657
 DB 598 SGADFYGABGEIYFKPTPRYRGVSGDYVRGLKNLPSLPGREDAYGNRPFFIAQDDQNA 657
 QY 658 RVPAAELGPHLKASLTDRLDANLDYRVFAQNKARLYETRTTCHHMLNLGANYRNRTRYG 717
 DB 658 RVPAAELGPHLKASLTDRLDANLDYRVFAQNKARLYETRTTCHHMLNLGANYRNRTRYG 717
 QY 718 EWNVYKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758
 DB 718 EWNVYKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758

RESULT 7

AAAY74564
 XX AAAY74564 standard; Protein; 758 AA.
 AC AAAY74564;
 XX 21-MAR-2000 (first entry)
 DT
 XX *Neisseria gonorrhoeae* ORF 149 protein sequence SEQ ID NO:602.
 DE
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS *Neisseria gonorrhoeae*.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0098994.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PA
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-PSDB; AA253326.
 XX
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 421-422; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA255941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA25437 to AA25476 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 758 AA;
 SQ Query Match 92.9%; Score 3751; DB 21; Length 758;
 Best Local Similarity 93.3%; Pred. No. 0;
 Matches 710; Conservative 13; Mismatches 32; Indels 6; Gaps 3;
 QY 1 MAQTTLKPIVLSILLINTPLLAQAHEHQSVGLTVTVVVKSRPRATSGLLHTSTASDKI 60
 DB 1 MAQTTLKPIVLSILLINTPLLAQAHEHQSVGLTVTVVVKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGASAPVIRGQTRRIKVLNHHGETGDMADF 120
 DB 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGASAPVIRGQTRRIKVLNHHGETGDMADF 120

QY 121 SPDHAIMVDTALSSQVEILLRGPVTLIYSSGNVAGLVADVGKIPKMPENGVSGLGLRL 180
 Db 121 SPDHAIMVDTALSSQVEILLRGPVTLIYSSGNVAGLVADVGKIPKMPENGVSGLGLRL 180
 QY 181 SGNLEKLTSGGINIGLGNFVLHTEGLYRKSVDVAPRYRNKRLPDSPPRFANGQRA 240
 Db 181 SGNLEKLTSGGINIGLGNFVLHTEGLYRKSVDVAPRYRNKRLPDSPPRFANGQRA 240
 QY 241 V-LGW--RRFVRRRTYSDRRDQVGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE 297
 Db 241 V-LGW--RRFVRRRTYSDRRDQVGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE 297
 QY 298 EDVDYDNPGLSCGFHDDDAHAHANGKPMWIDLRNRYELRAEWKQPPFGFPAALRVHLNR 357
 Db 298 EDVDYDNPGLSCGFHDDDAHAHANGKPMWIDLRNRYELRAEWKQPPFGFPAALRVHLNR 357
 QY 358 NDYHDEKAGDAVENFNQOTVARIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVK 417
 Db 358 NDYHDEKAGDAVENFNQOTVARIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVK 417
 QY 418 QPMLLDNKVQHYSGFFGVEQANWDFTLGGYRVEKOKASIRYDKALIDRENYKQPLPDL 477
 Db 418 QPMLLDNKVQHYSGFFGVEQANWDFTLGGYRVEKOKASIRYDKALIDRENYKQPLPDL 477
 QY 478 GAHQRTARSPALSGNRYFTPOHKLSTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 Db 478 GAHQRTARSPALSGNRYFTPOHKLSTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 QY 538 KERSNNIELALGYEGDRWQVNLALYRNFNGYIYAQTLDNGRGPKSIEDSEMKLVRYNQ 597
 Db 538 KERSNNIELALGYEGDRWQVNLALYRNFNGYIYAQTLDNGRGPKSIEDSEMKLVRYNQ 597
 QY 598 SGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGBEDAGNRPFTAQDONAP 657
 Db 598 SGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGBEDAGNRPFTAQDONAP 657
 QY 658 RVPARLGFHLKASLTDRIANDLYRFAQNKLAETRTFQHHMLNLGANYRNRTRYG 717
 Db 658 RVPARLGFHLKASLTDRIANDLYRFAQNKLAETRTFQHHMLNLGANYRNRTRYG 717
 QY 718 EWNWYVKAADNLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 758
 Db 718 EWNWYVKAADNLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 758

RESULT 8
 AAY74556
 ID AAY74556 standard; Protein; 735 AA.
 AC AAY74556;
 DT 21-MAR-2000 (first entry)
 DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:586.
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 XX antibacterial; gene therapy.
 XX Neisseria meningitidis.
 XX WO9957280-A2.
 PD 11-NOV-1999.
 XX 30-APR-1999; 99WO-US09346.
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 XX (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX MPI; 2000-062150/05.
 DR N-FSDB; AAZ53318.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 2; Page 413; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 735 AA;
 SQ Query Match 92.0%; Score 3714; DB 21; Length 735;
 Best Local Similarity 95.3%; Pred. No. 0;
 Matches 702; Conservative 10; Mismatches 19; Indels 6; Gaps 3;
 QY 25 HETQSVGLEFVTVVGVKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIH 84
 Db 2 HKTBSVDLETVSVVGVKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIH 61
 QY 85 ASQYGGGASAPVIRGQTRRIKVLNHHGETGMADFSFDHAIMVDTALSSQVEILLRGPVT 144
 Db 62 ASQYGGGASAPVIRGQTRRIKVLNHHGETGMADFSFDHAIMVDTALSSQVEILLRGPVT 121
 QY 145 LLYSSGNVAGLVADVGKIPKMPENGVSGLGLRLSGNLEKLTSGGINIGLGNFVLH 204
 Db 122 LLYSSGNVAGLVADVGKIPKMPENGVSGLGLRLSGNLEKLTSGGINIGLGNFVLH 181
 QY 205 TEGLYRKSVDVAPRYRNKRLPDSPPRFANGQRAV-LGW--RRFVRRRTYSDRRDQYG 261
 Db 182 TEGLYRKSVDVAPRYRNKRLPDSPPRFANGQRAV-LGW--RRFVRRRTYSDRRDQYG 238
 QY 262 LPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHA 321
 Db 239 LPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHA 298
 QY 322 HNGKPMWIDLRNRYELRAEWKQPPFGFPAALRVHLNRNDYHDEKAGDAVENFNQOTNA 381
 Db 299 HNGKPMWIDLRNRYELRAEWKQPPFGFPAALRVHLNRNDYHDEKAGDAVENFNQOTNA 358
 QY 382 RIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVKQPMLLDNKQHYSGFFGVEQANWDN 441
 Db 359 RIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVKQPMLLDNKQHYSGFFGVEQANWDN 418
 QY 442 FTLEGGYRVEKOKASIRYDKALIDRENYKQPLPDLGAHQRTARSPALSGNRYFTPOHKL 501
 Db 419 FTLEGGYRVEKOKASIRYDKALIDRENYKQPLPDLGAHQRTARSPALSGNRYFTPOHKL 478
 QY 502 SLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLNERSNNIELALGYEGDRWQVNLAL 561
 Db 479 SLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLNERSNNIELALGYEGDRWQVNLAL 538

562 YRNRFGNYIAQTINDRGPKSIEDDSEMKLVRYNOSGADFYAGSGEIIYFKPTPRYRIGV 621
 539 YRNRFGNYIAQTINDRGPKSIEDDSEMKLVRYNOSGADFYAGSGEIIYFKPTPRYRIGV 598
 622 SGDVVRGLKNLPSLPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLD 681
 599 SGDVVRGLKNLPSLPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLD 658
 682 YRVFAQNKARVETRTPGHHMLNLGANYRNTRYGHWNYVADNLLNOSVVAHSSFLS 741
 659 YRVFAQNKARVETRTPGHHMLNLGANYRNTRYGHWNYVADNLLNOSVVAHSSFLS 718
 742 DTPQGRSFTGGVNVKF 758
 719 DTPQGRSFTGGVNVKF 735
 RESULT 9
 ABP80923 standard; Protein; 728 AA.
 XX ABP80923;
 AC ABP80923;
 DT 07-MAR-2003 (first entry)
 XX N. gonorrhoeae amino acid sequence SEQ ID 8376.
 DE Antibacterial; infection; vaccine; gene therapy.
 KW Neisseria gonorrhoeae.
 OS Neisseria gonorrhoeae.
 XX WO200279243-A2.
 FN 10-OCT-2002.
 PD 12-FEB-2002; 2002WO-IB02069.
 PF 12-FEB-2001; 2001GB-0003424.
 PR (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizza M, Massignani V, Monaci E;
 XX WPI: 2003-058415/05.
 DR N-PSDB; ABZ41893.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 XX Disclosure; Page 799; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX Sequence 728 AA;
 SQ
 Query Match 90.3%; Score 3644; DB 24; Length 728;
 Best Local Similarity 93.8%; Pred. No. 0;
 Matches 686; Conservative 11; Mismatches 28; Indels 6; Gaps 3;
 31 VLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIHASQYG 90
 1 VLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIHASQYG 60
 91 GASAPVIRGQTRRIKVLNHHGEGTGMADFPDHAIMVDTALSOQVIRLGPVTLTYSSG 150

61 GASAPVIRGQTRRIKVLNHHGEGTGMADFPDHAIMVDTALSOQVIRLGPVTLTYSSG 120
 151 NVAGLVADVADGKIPEKMPENGVSGELGLRLSSGNLEKLTSGGINIGLQKPFVLTGELYR 210
 121 NVAGLVADVADGKIPEKMPENGVSGEAGLRSLSSGNLEKLTSGINIGLQKPFVLTGELYR 180
 211 KSGDYAVPRYRNLRKLPDSRRFANGQHRV-LGW--RKPFYRRTYSRRRDQYGLPAHSH 267
 181 KSGDYAVPRYRNLRKLPDS---HADSGTSGISGLSVGKGFAGYSDRRDRYGLPAHSH 237
 268 EYDDCHADIIWOKSLINKRYLQLYPHLLTBEVDVYDNPGLSCGHDDDDAHAAHAKPKW 327
 238 EYDDCHADIIWOKSLINKRYLQLYPHLLTBEVDVYDNPGLSCGHDDGGAHAHTNCKPW 297
 328 IDLRNKRVELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNTONARIELRH 387
 298 IDLRNKRVELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNTONARIELRH 357
 388 QFIGRLKSGWGVQYLGOKSSALSATSSEAVKQPMLLDNKVHYSFFGVEQANWDFTLGG 447
 358 QFIGRLKSGWGVQYLGOKSSALSATSSEAVKQPMLLDNKVHYSFFGVEQANWDFTLGG 417
 448 VRVEKOKASIRYDKALIDRENYKQPLDLGAHQRTARSFALSGNWTFTPOHKLSTASH 507
 418 VRVEKOKASIRYDKALIDRENYKQPLDLGAHQRTARSFALSGNWTFTPOHKLSTASH 477
 508 QERLPSTQELYAHGKHVATNTFEVGNKHLNKNERNIELALGYEGDRWQYNLALYRNF 567
 478 QERLPSTQELYAHGKHVATNTFEVGNKHLNKNERNIELALGYEGDRWQYNLALYRNF 537
 568 NYIYAQTINDRGPKSIEDDSEMKLVRYNOSGADFYAGSGEIIYFKPTPRYRIGVSGDYVR 627
 538 NYIYAQTINDRGPKSIEDDSEMKLVRYNOSGADFYAGSGEIIYFKPTPRYRIGVSGDYVR 597
 628 GRLKNLPSLPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVFA 687
 598 GRLKNLPSLPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVFA 657
 688 QNKLYRTRTPGHHMLNLGANYRNTRYGHWNYVADNLLNOSVVAHSSFLSDTPMG 747
 658 QNKLYRTRTPGHHMLNLGANYRNTRYGHWNYVADNLLNOSVVAHSSFLSDTPMG 717
 748 RSFTGGVNVKF 758
 718 RSFTGGVNVKF 728
 RESULT 10
 AAY74562
 ID AAY74562 standard; Protein; 339 AA.
 XX AAY74562;
 AC AAY74562;
 DT 21-MAR-2000 (first entry)
 XX Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:598.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX Neisseria meningitidis.
 OS Neisseria meningitidis.
 XX WO957280-A2.
 PN 11-NOV-1999.
 PD 30-APR-1999; 99WO-US09346.
 XX 31-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.


```

PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53324.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 419- ; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ54253 to AAZ55941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX Sequence 339 AA;
XX Query Match 44.8%; Score 1808; DB 21; Length 339;
XX Best Local Similarity 99.1%; Pred. No. 8.4e-157;
XX Matches 336; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 420 MLLDNKVQHYSPFGVEQANWDFTEGGVVRVEKOKASIRYDKALIDRENYKQPLDLGA 479
Db 1 MLLDNKVQHYSPFGVEQANWDFTEGGVVRVEKOKASIQYDKALIDRENYNHPDLGA 60
QY 480 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 539
Db 61 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 120
QY 540 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQSG 599
Db 121 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQSG 180
QY 600 ADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFTAQDDONAPRV 659
Db 181 ADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFTAQDDONAPRV 240
QY 660 PAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLCANYRNTRYGEW 719
Db 241 PAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLCANYRNTRYGEW 300
QY 720 NWYKADNLLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 758
Db 301 NWYKADNLLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 339
RESULT 11
AAZ74563
ID AAY74563 standard; Protein; 339 AA.
XX AAZ74563;
XX 21-WAR-2000 (first entry)
XX

```

```

DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:600.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX antibacterial; gene therapy.
XX OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53325.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 420; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ54253 to AAZ55941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX Sequence 339 AA;
XX Query Match 44.4%; Score 1791; DB 21; Length 339;
XX Best Local Similarity 98.5%; Pred. No. 3.1e-155;
XX Matches 334; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 420 MLLDNKVQHYSPFGVEQANWDFTEGGVVRVEKOKASIRYDKALIDRENYKQPLDLGA 479
Db 1 MLLDNKVQHYSPFGVEQANWDFTEGGVVRVEKOKASIRYDKALIDRENYNHPDLGA 60
QY 480 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 539
Db 61 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 120
QY 540 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQSG 599
Db 121 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQSG 180
QY 600 ADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFTAQDDONAPRV 659
Db 181 ADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFTAQDDONAPRV 240

```


QY 660 PAARLGFHLKASLDRIDANLDYRVFAQNKLYARVETTPGHHMLNLGANYRNRTRYGEW 719
 DB 241 PAARLGVHLKASLDRIDANLDYRVFAQNKLYARVETTPGHHMLNLGANYRNRTRYGEW 300
 QY 720 NWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 DB 301 NWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 339

RESULT 12

ID AAY74561 standard; Protein; 339 AA.
 XX AAY74561;
 DT 21-MAR-2000 (first entry)
 XX Neisseria gonorrhoeae ORF 149 protein sequence SEQ ID NO:596.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX Neisseria gonorrhoeae.
 OS WO9957280-A2.
 PN 11-NOV-1999.
 PD 30-APR-1999; 99WO-US09346.
 PF 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0098062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.

PA Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 DR N-PSDB; AA253323.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 2; Page 418; 1453pp; English.
 XX AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA254941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254941 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 339 AA;
 SQ Query Match 43.7%; Score 1763; DB 21; Length 339;

Best Local Similarity 96.2%; Pred. No. 1.1e-152;
 Matches 326; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
 QY 420 MLIDNKKVQHSYFQVQANWDNFTLEGGRVVEKQKASTRYDKALIDRENYKQPLDLAGA 479
 DB 1 MLIDNNVRHSYFQVQANWDNFTLEGGRVVEKQKASTRYDKALIDRENYKQPLDLAGA 60
 QY 480 HROTARSPALSGNWYFTPOHKLSTLASHOERLPSTQELYAHGKHVATNTFEVGNKHLNKE 539
 DB 61 HROTARSPALSGNWYFTPHHKLSTLASHOERLPSTQELYAHGKHVATNTFEVGNKHLNKE 120
 QY 540 RSNNIELALGYEGDRQYNLALYRNRFGNYIYAQTINDGRGPKSIEDDSEMKLVRYNOSG 599
 DB 121 RSNNIELALGYEGDRQYNLALYRNRFGNYIYAQTINDGRGPKSIEDDSEMKLVRYNOSG 180
 QY 600 ADPYGAEGEIVFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAPEV 659
 DB 181 ADPYGAEGEIVFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAPEV 240
 QY 660 PAARLGFHLKASLDRIDANLDYRVFAQNKLYARVETTPGHHMLNLGANYRNRTRYGEW 719
 DB 241 PAARLGFHLKASLDRIDANLDYRVFAQNKLYARVETTPGHHMLNLGANYRNRTRYGEW 300
 QY 720 NWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 DB 301 NWYVKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 339

RESULT 13

AAO17586
 ID AAO17586 standard; Protein; 814 AA.
 XX AAO17586;
 DT 19-JUL-2002 (first entry)
 DE M catarrhalis MCA102062 protein SEQ ID NO: 52.
 XX Moraxella, vaccine; respiratory tract infection; antiinflammatory;
 KW auditory; antibacterial; otitis media; sinusitis; pneumonia.
 XX Moraxella catarrhalis.
 OS WO200218595-A2.
 PN 07-MAR-2002.
 PD 28-AUG-2001; 2001WO-CA01221.
 PF 28-AUG-2000; 2000US-228294P.
 PR 28-AUG-2000; 2000US-228295P.
 PR 28-AUG-2000; 2000US-228296P.
 PR 29-AUG-2000; 2000US-228438P.
 PR 29-AUG-2000; 2000US-228439P.
 PR 29-AUG-2000; 2000US-228440P.
 PR 29-AUG-2000; 2000US-228441P.
 PR 29-AUG-2000; 2000US-228442P.
 PR 29-AUG-2000; 2000US-228443P.
 PR 29-AUG-2000; 2000US-228512P.
 PR 29-AUG-2000; 2000US-228513P.
 PR 29-AUG-2000; 2000US-228742P.
 PR 29-AUG-2000; 2000US-228773P.
 PR 01-SEP-2000; 2000US-229465P.
 PR 01-SEP-2000; 2000US-229474P.
 PR 01-SEP-2000; 2000US-229475P.
 PR 01-SEP-2000; 2000US-229476P.
 PR 05-SEP-2000; 2000US-229740P.
 PR 05-SEP-2000; 2000US-229803P.
 PR 05-SEP-2000; 2000US-229804P.
 PR 05-SEP-2000; 2000US-229805P.
 PR 05-SEP-2000; 2000US-229806P.
 PR 05-SEP-2000; 2000US-229809P.
 PR 05-SEP-2000; 2000US-229811P.

PR 06-SEP-2000; 2000US-230214P.
 PR 06-SEP-2000; 2000US-230250P.
 PR 06-SEP-2000; 2000US-230252P.
 XX (AVET) AVENTIS PASTEUR LTD.
 XX Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
 XX WPI; 2002-401721/43.
 XX N-PSDB; AAL46516.
 XX Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX Claim 28; Fig 51; 277pp; English.
 XX The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a protein of the invention.
 XX Sequence 814 AA;
 SQ
 Query Match 39.2%; Score 1580.5; DB 23; Length 814;
 Best Local Similarity 41.2%; Pred. No. 2.6e-135;
 Matches 336; Conservative 141; Mismatches 270; Indels 69; Gaps 16;
 QY 7 KPIVLSIL-LINTPLLAQAH-ETQSVGLETVTVVGVKSPRATSGLLHTSTASDKIISGD 64
 DB 4 KPLACAILATFSPMLAEANLKDRPTVLDGVSTSLADQNTFEGVNHSTVSGITVSKE 63
 QY 65 TLQKAVNLGALDGVPGIHAISQYGGASAPVIRGQTRIKVLNHHGETGDMADSPDH 124
 DB 64 QLOQRATTLGALAGELGVHSHFGGASAPVIRGQTRIKVLNHHGETGDMADSPDH 123
 QY 125 AIMVDTALSOQVEILRGPTVLLYSSGNVAGVADVADGKIPEKMPENGVSGLGLRLSSGN 184
 DB 124 AIAVDVTLAKOVELVIRGSGALLYASGNSAGVNVVDDKIPEKLPFSK-LQGDVTVRLSSAN 182
 QY 185 LEKLTSGGINILGLGNFVLHTEGLYRKSQGVAVPRY-----RNL 223
 DB 183 REKLITASAEAPLGEHVAVRVAGLSKQADYKTRFDRHVFNKXKEDNTQPEFIYKDTL 242
 QY 224 KRLPDSPPRRFANGQRAVLG--W--RKFPYRTVSDRRDQVGLPAHSHVEDDCHADLIHQ 279
 DB 243 KHLFDSHAK-----SNAGTGLVSWVGNQGLGASVSLRRDKYGLPNHSHYEYECVHGISQ 298
 QY 280 KSLINKRYLQLYPHLLTBEDVDYDNPGLSCGFHDD-----DDAHAAH-----NGKFWID 329
 DB 299 SALQYKFLVRLYPLFMENDDLLEFDNAGLECHTHDDHDDHDDHDDHDDHDDHDDHDDH 358
 QY 330 LNKRYELRAEWKQPPFGFEALRVLNNDVHHDEKAGDAVENFNNTQNAIELRHQP 389
 DB 359 LMKRYDVQGOINAPFAGIDIRASMGKVDVHHDEIDGGEKTSFFDNQANVRLEASHTP 418
 QY 390 I-----GRKLSGWQVYLQKSSAL-----SATSAVQKPMLLNKKVQHYSGFEGEQAN-W 439
 DB 419 IHTPWKSGVFGVGYLTSKNSGLVPPRYEDGNGKQDTQNTILHNKNTKGTGVFFEEYKPN 478
 QY 440 DNFTLEGGVRVEKQKASIRYDKALI-----DRENYKQPLPD-----LGAHQRTAR 485
 DB 479 DKLTVDAAARTEKQITMDYDKDAIYQSLNLGLATAHEPDIRFKRLDSDGLNPKQTAR 538
 QY 486 SFALSGNWFYFQHKLSLTASHQELPSTOBYLHAGHVAHNTFVGNKHLNKSNNIE 545
 DB 539 SYAVETHLQLTQPKKLSLNLHSHQELPNAQBYLHAGHVAHNTFVGNKHLNKSNNID 598
 QY 546 LALYEGDRWQVNLALYRNRFGNIYIAQTLND-GRGPKSIEDDSEMKLVRYNQSADFYG 604
 DB 599 LGLTFQGGKNDYRLGGVHYDFNVFLQTLQSYQKGLRGHDKDLKTARYEQAAAFYG 658

QY 605 AEGEIVFKPTPRYRIGVSGDYYVRGKLNKPSLPGREDAYGNRPFIQAQDQNAIPVPAARL 664
 DB 659 FQVNIQYQINDVYHVALFGDVIKGLTNLPDKKGRTDAYGNRPFLIKQDSDHTPLPPLRL 718
 QY 665 GFHLKASLTDRIDANLDIYRVFAQNKIARYTRTPGHMLNLGANYRNTY--GEWNWY 722
 DB 719 GMKLTANTVANWSGFLRYRHTFKQDKLANFERPTPAHNLVNLGLNYQHKPSHQAGSVQVF 778
 QY 723 VKADNLLNQSYYAHSSFLSDTPOMGSRFTGGVNVKF 758
 DB 779 FNANLLNDKVFPAHETFFDFDMPQMGSRNFMILGANFKF 814
 RESULT 14
 AAB19945
 ID AAB19945 standard; Protein; 818 AA.
 XX AAB19945;
 AC AAB19945;
 XX 19-MAR-2001 (first entry)
 DT Moraxella catarrhalis BASB107.
 XX Moraxella catarrhalis BASB107.
 DE
 XX
 KW BASB107; infection; pneumonia; otitis media; sinusitis;
 KW therapy; diagnosis; vaccine; genetic immunization; antibacterial;
 KW antibiotic; receptor.
 XX
 OS Moraxella catarrhalis.
 OS WO2000071724-A2.
 PN 30-NOV-2000.
 PD 18-MAY-2000; 2000WO-EP04618.
 PF 24-MAY-1999; 99GB-0012038.
 PR 24-MAY-1999; 99GB-0012040.
 PR 01-JUN-1999; 99GB-0012674.
 PR 01-JUN-1999; 99GB-0012705.
 PR 02-JUN-1999; 99GB-0012838.
 PR 08-JUN-1999; 99GB-0013354.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA Thonnard J;
 XX WPI; 2001-025166/03.
 DR N-PSDB; AAA89216.
 XX New BASB103-108 polypeptides isolated from Moraxella catarrhalis
 PT bacterium, useful for diagnosing and producing vaccines against
 PT bacterial infections such as otitis media and pneumonia
 XX Claim 1; Page 74-75; 79pp; English.
 PS The present sequence is that of Moraxella catarrhalis ATCC 43617
 CC BASB107, a novel protein that shows amino acid sequence homology
 CC to Escherichia coli PhuE receptor precursor (outer membrane receptor
 CC for Fe(III)-coprogen, Fe(III)-ferritoxamine-B and Fe(III)-rhodotorulic
 CC acid). The invention provides M. catarrhalis BASB103-108
 CC polypeptides (see AAB19941-46) and polynucleotides (see AAA89212-17),
 CC and methods for producing the polypeptides by recombinant methods.
 CC Claimed vaccine compositions comprise a BASB103-108 polypeptide, or
 CC a BASB103-108 polynucleotide for genetic immunization. A claimed
 CC method for diagnosing a M. catarrhalis infection involves identifying
 CC a BASB103-108 polypeptide, or an antibody immunospecific to a
 CC BASB103-108 polypeptide, in a biological sample. Compositions
 CC comprising an immunologically effective amount of a BASB103-108
 CC polypeptide, or a polynucleotide encoding it, are used to generate
 CC an immune response in an animal. An antibody directed against a
 CC BASB103-108 polypeptide can be used to treat humans with M.
 CC catarrhalis disease. The polypeptides are also used as research
 CC reagents for the discovery of therapeutics and diagnostics useful

RESULT 15

```
QY 177 GLRLSSGNLEKLTSGGINIGLGNFVHLHTGLYRKSGDYAVPRY---RNLRPLDPSRRF 233
Db 181 MLRYNTNNHEKLATAGVSGVGRDRIAVRVEGLKREADDYQVPHFQADRLMDYVPGS---- 236
QY 234 ANQGHRAVLG----WRKRFRYRRYSRRRDYGLPAHSHSYDDCHADIIWOKSLINKRYLQ 289
Db 237 ANNSTVGMIGVSYIHDNGHIGASYSRKRDRYGIPIGHIHCDQSREHFIKWHNITKSNYLP 296
QY 290 LYPHLLTSEDDVDNPGSLGCPHDDDDAHAAHNG-----KPWIDLENKYELEAEW 341
Db 297 IYPLHMESSDID--DNPHTHCRNHEHDIHGHNPTGVPINHEHSPWIDMKTNRDYIRGEV 355
QY 342 KOPFPGEALRVLHNLNDYHDEK--AGDA-----VENFFNNOTQNA 381
Db 356 YRPIQLGDKIKLSLTADYIYHDEKADGNEQDPNNHKPSEDDTTVDKCHASSITFKGVNG 415
QY 382 RIELRHQPTGRKSGVQVGLQKQSALSALSA-----TSEAVKQP-----MLLD 423
Db 416 RLELYHTPTKRLSGVLGIGYGTQKSAAGEAYLPSYFQSEAEWQAQSONINQVRPYLLVP 475
QY 424 NKVQHSYFQVEQANNDNFTLEGGVVEKOKASIRYDKALIDRENY-----KQP-LP 475
Db 476 NTKSLGIGFGLQKLNQMTFKVAMRHERKQTEIYDQHLHDHALQYFLSKAQLKAPDHP 535
QY 476 DLGAHQRTARSFALSGNWYFTPOHKLSTASHOERLPSTOELYAHGKHVATNTFEVGNKH 535
Db 536 DLTTYQHATSYAGSALWDITPNHRLSLTYSHNERIPSPMELYQGHLATSSFEHGNKN 595
QY 536 LNKERSNNIELALGYEGDRWQYNLALRYRNFQNYIYAQTLNDGRGPKSIEDDSEMKLVR 595
Db 596 LVKEKSDNYELGPMHTADKVSYKASYYSNFDNYIFNET-----IAKEGNLYIRRY 646
QY 596 NOSGADFYGAEGRIYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAODQ- 654
Db 647 NQTTAKFYVEGSLTYQPNANHSVMFQGMVQSKIGALSIDIKGL-VYAGRKWVYFDDDI 705
QY 655 ----- 654
Db 706 KDMTVDNDGDYDADGGLTCAKLTPEQWQINDNDCSTTINVYKNGTITTSGBEDYDLAR 765
QY 655 ---NAPRVPAAIRLGHKLSLDRIDANLDLYRVFAQNKLA----- 692
Db 766 NPTYAPRVPSPRLGIRWQGYFGDHWASANAENHVAQNKVATSTVAIKPQKQEGCQRH 825
QY 693 -----RY--ETRTQGHMLNLGANYRNRTRGYENWVYKADNLNL 730
Db 826 ESHCRISDYGSDNNPLMQPRYITENKTAGYNLLNVLGDDYNNAYRNVDTLSIRANLLN 885
QY 731 QSVYAHSSFLSDPTQMGSRFTGQVNVKF 758
Db 886 EQIYIHNSFLFPQWQGRNLTGLTAKF 913

RESULT 16
AA017585
ID AA017585 standard; Protein; 913 AA.
XX
AC AA017585;
XX
DT 19-JUL-2002 (first entry)
XX
DE M catarrhalis MCA101976 protein SEQ ID NO: 50.
XX
KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;
XX auditory; antibacterial; otitis media; sinusitis; pneumonia.
OS Moraxella catarrhalis.
XX
PN WO200218595-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-CA01221.
```

```
XX 28-AUG-2000; 2000US-228294P.
PR 28-AUG-2000; 2000US-228295P.
PR 28-AUG-2000; 2000US-228296P.
PR 29-AUG-2000; 2000US-228438P.
PR 29-AUG-2000; 2000US-228439P.
PR 29-AUG-2000; 2000US-228440P.
PR 29-AUG-2000; 2000US-228441P.
PR 29-AUG-2000; 2000US-228442P.
PR 29-AUG-2000; 2000US-228443P.
PR 29-AUG-2000; 2000US-228511P.
PR 29-AUG-2000; 2000US-228512P.
PR 29-AUG-2000; 2000US-228742P.
PR 29-AUG-2000; 2000US-228773P.
PR 01-SEP-2000; 2000US-229465P.
PR 01-SEP-2000; 2000US-229474P.
PR 01-SEP-2000; 2000US-229475P.
PR 01-SEP-2000; 2000US-229478P.
PR 05-SEP-2000; 2000US-229740P.
PR 05-SEP-2000; 2000US-229803P.
PR 05-SEP-2000; 2000US-229804P.
PR 05-SEP-2000; 2000US-229805P.
PR 05-SEP-2000; 2000US-229806P.
PR 05-SEP-2000; 2000US-229809P.
PR 05-SEP-2000; 2000US-229811P.
PR 06-SEP-2000; 2000US-230214P.
PR 06-SEP-2000; 2000US-230250P.
PR 06-SEP-2000; 2000US-230252P.
XX
PA (AVET ) AVENTIS PASTEUR LTD.
XX
PI Loomore S, Wang J, Bradley B, Ochs M, Yang Y;
XX
XX WPI; 2002-401721/43.
DR N-PSDB; AAL46517.
XX
PT Moraxella polypeptide and polynucleotides useful as vaccine for
PT immunizing a host e.g. humans against disease e.g. otitis media,
PT pneumonia, caused by infection of the bacteria
XX
XX Claim 28; Fig 49; 277pp; English.
XX
XX The present invention provides the protein and coding sequences of
XX proteins from Moraxella catarrhalis. These can be used to produce
XX vaccines which protect against M. catarrhalis infection, which can cause
XX otitis media, respiratory infection, sinusitis, and pneumonia. The
XX present sequence is a protein of the invention.
XX
XX Sequence 913 AA;
XX
XX Query Match 29.8%; Score 1202; DB 23; Length 913;
XX Best Local Similarity 31.4%; Pred. No. 1.6e-100;
XX Matches 291; Conservative 142; Mismatches 301; Indels 194; Gaps 20;
QY 7 KPVLISILLINTPLLAQAHAHETQSGVLEVT---VVGSRPRATSGLLHTST--ASDKI 60
Db 4 KPLVCA---ISATFAMPADVADNTKLGEETTTIKGLVSSQTNQNTGFSVNSDKSQSSDLT 60
QY 61 ISGDTLRQKAVNLDALDVGPIHASQVGGGASAPVIRGOTGRRIKVLNHHGETGMDADF 120
Db 61 LSKDKLKYRSATLGNALSGELGHNPGGSSAPVVRGEGVRLKILQNGTVDIVDSI 120
QY 121 SPDHAMVDITALSQQVEILRGVPTLLYSNGVAGLVADVADKIPKMPENG----VSGEL 176
Db 121 SPDHVAVDTLLASKVELVRGADTLLYGLASPAGVINVDRIIPNEMPSGAHDKIEGT 180
QY 177 GLRLSSGNLEKLTSGGINIGLGNFVHLHTGLYRKSGDYAVPRY---RNLRPLDPSRRF 233
Db 181 MLRYNTNNHEKLATAGVSGVGRDRIAVRVEGLKREADDYQVPHFQADRLMDYVPGS---- 236
QY 234 ANQGHRAVLG---WRKRFRYRRYSRRRDYGLPAHSHSYDDCHADIIWOKSLINKRYLQ 289
Db 237 ANNSTVGMIGVSYIHDNGHIGASYSRKRDRYGIPIGHIHCDQSREHFIKWHNITKSNYLP 296
```


XX WO9631618-A1.
 XX 10-OCT-1996.
 XX 08-APR-1996; 96WO-US04774.
 XX 07-APR-1995; 95US-0418964.
 XX (UNNC-) UNIV NORTH CAROLINA.
 XX Beucher M, Sparling PF;
 XX WPI; 1996-465038/46.
 XX N-PSDB; AAT42128.
 XX DNA encoding *Neisseria* iron-regulated outer membrane protein FrpB -
 XX useful for producing vaccine to protect mammals against *N.*
 XX gonorrhoeae and *N. meningitidis* infection
 XX Claim 1; Fig 10; 68pp; English.
 XX This sequence represents a novel iron-regulated outer-membrane protein
 XX FrpB which was isolated from the gonococcal microorganism *Neisseria*
 XX gonorrhoeae strain FA1090. This gene is also homologous to one found in
 XX *Neisseria meningitidis*. The FrpB protein is surface exposed and
 XX immunogenic and is predicted to be useful as a vaccine because of its
 XX surface exposure, partial antigenic conservation and susceptibility to
 XX attack by bacterial antibodies. Such a vaccine will be of use against
 XX infection in mammals by *N. gonorrhoeae* and *N. meningitidis*. The protein
 XX and its fragments will also be useful in the diagnosis of *Neisseria*
 XX type infections.
 XX
 XX Sequence 713 AA;
 Query Match 5.4%; Score 216; DB 17; Length 713;
 Best Local Similarity 21.6%; Pred. No. 2.7e-10;
 Matches 187; Conservative 103; Mismatches 298; Indels 278; Gaps 47;
 QY 11 LSLLINPL-LAQAHETEQQVGLTETVVVGVKSPRATSGLLHTSTASDKIISGDTLRLQK 69
 DB 8 LSLLSLTLAAGFAHAENANVALDTTVKGDQ-----GSKIRTIV---TLQOK 55
 QY 70 ----AVNLGDALDGVPGTHAQVGGASAPVI-----RGOTGREIKV-----LNHH 111
 DB 56 DESTATDRELKKEPSI---DFGGNGTSQFLIRGMGQNSVDIKVDNAYSDSQILYHQ 112
 QY 112 GETGDMADFPDHAIMVDLTALSOQVEILRLGPVTLTYSGSNVAGLVVDVADGKIPKEMPE-- 169
 DB 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDQAQ 153
 QY 170 ---NGVSGELGLRLSSGNLEKLTSGINIG---LGK-NFVLHTEGLY---RKSGDYAV 217
 DB 154 DLLKGLDKNWGVRLNSG---FAGNGVSYGASVFGKEGNF---DGLFSVNRNDEKDYEA 206
 QY 218 PR-YRNL---KRLPDSP-----RRFANGQHRVAVLGMWRKRFVR--RTYSDRRDQY 260
 DB 207 GKGFNVNGKTVPVYALDKRSYLAKIGTTFDGDHRLVLSHMKDQHRGINTV---REEF 263
 QY 261 GLPAHSHEYDCHADIIWQSLINKRYLQPLPHLLTEEDVDYDNPGLSCGHDDDDAHAAH 320
 DB 264 AVYGEN-----SRITIKRQAPAYRE-TTQSNITNLAYTGKDLGFVEKLDANAY 309
 QY 321 AHNGKFWIDLRNKRVELRAE---WKQPFGEALRVHLNRNDYHDEKAGDAVENFFNQ 377
 DB 310 V-----LEKRYSAADKNGYAGNVKGNPHRTIATGNNFNFDSLAE-----Q 353
 QY 378 T-QNARIELRHQ---PIGLKSGWGVQYLQKSSAL-----SATSEAVKQPMLLDNKVQ 427
 DB 354 TLLKYGINVRHQEIKPQAFNLNSQFKIE---DKKQATEEDKKKRENEKIAKAYRLTPTK 410
 QY 428 HYSFPGVEQAN-WDNFTLEGGVVRVEKQKASITRYDKALIDRENYKQPLPDGLAHQRTARS 486

DB 411 TDTGAY1EAIHIDGFTLTGGLRYDRFKVKTGDK-----TVSS 449
 QY 487 FALSGNW--YFTPOKLSLTASHQ--ERLPTQE-LYAHGKH---VATNTFEVGNKHLN 537
 DB 450 SSLNPSFGVWQPREHWSFSASHNVASRPLYDALQTHGKGIISADGT-----K 501
 QY 538 KERSNNIELALGYEGDRWOYNLALYR-----NRFNGYIYAOQLNDG----- 578
 DB 502 AERARTEIGFNYNDGTFAANGSVFRQTIKDALANPQNRHDSVAVREAVNAGYIKHGYE 561
 QY 579 -----FGPKSIEDDSEMKLVRYNQSGADFYGAB-GEIY-----FKP 613
 DB 562 LGASYRTGGLTAKVGVSRKPRFYDTHPKKLLSANPE---FGATGTWTASLAYRKF- 616
 QY 614 TPRYRIGVSGDYVRGRLKNLPSLPQREDAYGNRPPIAQQDQONAPRVPAARLGFHLKASLT 673
 DB 617 NFNLEIGWRGRIV-----QKATGSILAAQKDRDKLENVVRQCGFVN----- 659
 QY 674 DRIDANLDYRVFAQNKLYARTETPTGHHMLNLGANYRNTRYGEWNYVVRADNLLNQSV 733
 DB 660 -----DVFANWKPLGKDTLN-----VNLSVN-----NVDFKPY 687
 QY 734 YAHSSFLSDT-PQMGKSFYGGVNVKVF 758
 DB 688 YPHSORWTNLTGVRDVRVLGVNYKF 713
 RESULT 19
 AAE05408
 ID AAE05408 standard; Protein; 713 AA.
 XX AAE05408;
 AC AAE05408;
 XX
 DT 24-SEP-2001 (first entry)
 XX
 DE *N. gonorrhoeae* or *N. meningitidis* strain FA1090 FrpB protein.
 XX
 KW Fe-regulated protein B; FrpB; outer-membrane protein; vaccine; infection;
 KW antibacterial.
 XX
 OS *Neisseria gonorrhoeae*.
 OS *Neisseria meningitidis*.
 XX
 PN US6265567-B1.
 XX
 PD 24-JUL-2001.
 XX
 PF 05-APR-1996; 96US-0628434.
 XX
 PR 07-APR-1995; 95US-0418964.
 XX
 PA (UNNC-) UNIV NORTH CAROLINA.
 XX
 PI Sparling PF, Beucher M;
 XX WPI; 2001-450739/48.
 DR N-PSDB; AAD10202.
 XX
 PT An isolated nucleic acid encoding a Fe-regulated protein B (FrpB),
 XX useful for production of a vaccine against infection by *Neisseria*
 XX gonorrhoeae or *N. meningitidis*.
 PS Claim 2; Column 33-38; 20pp; English.
 CC The present sequence is a Fe-regulated protein B (FrpB) isolated
 CC from *Neisseria gonorrhoeae* and *Neisseria meningitidis* (strain FA1090). The
 CC present DNA sequence also represents the combined nucleotide sequence
 CC from PUNCH319 and PUNCH325. The FrpB is a 70 kD major iron-regulated,
 CC outer-membrane protein common to *N. gonorrhoeae* or *N. meningitidis*.
 CC The FrpB has particular use as a vaccine that protects a mammal from
 CC *N. gonorrhoeae* or *N. meningitidis* infection. The FrpB may also be used
 CC to detect the presence of antibodies specific for *N. gonorrhoeae* or
 CC *N. meningitidis* in a sample.

XX Sequence 713 AA;
 Query Match 5.4%; Score 216; DB 22; Length 713;
 Best Local Similarity 21.6%; Pred. No. 2.7e-10;
 Matches 187; Conservative 103; Mismatches 298; Indels 278; Gaps 47;

QY 11 LSILLINTPL-LAQAHETEQSGLETVTVVVKSRPRATSGLLHTSTASDKLIISGDTLRQK 69
 DB 8 LSLLSLTAAAGFAHAENANVALDTVTVKGRQ-----GSKRTNIV---TLQOK 55

QY 70 ----AVNLGDALDGVPGIHASQYGGASAPVI-----RGOTGRRIKV-----LNHH 111
 DB 56 DESTATDMRELLKEEPSI---DFGGNGTSGFTLRGMGQNSVDIKVDNAYSQSILYHQ 112

QY 112 GETGDMADFPDHAIVDTALSOQVEILLRQPVTVLLYSSGNVAGLVADVADKIKEKPE-- 169
 DB 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDAQ 153

QY 170 ---NGVSGELGLRLSSGNLEKLTSGGINIG---LGK---NFVLHTEGLY---RKSGDYAV 217
 DB 154 DLLKGLDKXWVRLNSG---FAGNNGVSYGASVFGKEGNF---DGLFSYNRNDEKDYEA 206

QY 218 PR-YRNL---KRLPDSP-----RRFANGQHRVILGWRKPFYR---RTYSDRRQY 260
 DB 207 GKGFNRVNGGKTVPSYALDKRSYLAKIGTTFGDDHRIVLSHMKDQHRGIRTV---REEF 263

QY 261 GLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAH 320
 DB 264 AVGEN-----SRITIKRQAPAYRE---TTQSTNTNLAYTKDGLGFVEKLDANAY 309

QY 321 AHNGKPMIDLNRKRYELRAE---WKQPPGFPALRVHLNRNDYHDEKAGDAVENFNQ 377
 DB 310 V-----LEKKYSADDDKNGYAGNVKGNPNHTRATRCGMNFNDSRLAE-----Q 353

QY 378 T-QNARIELRHQ---PIGLKSGWGYOYLGQKSSAL-----SATSEAVKQPMLLDNKVQ 427
 DB 354 TLLKVCINVRHGEIKPQAFINSQFIE---DKDATEEDKKKRENEKIAKAYRLTNTPK 410

QY 428 HYSFFGVEQAN-WDNFTLEGGVVRVEKQASIRYDKALIDRENYYKQPLDILGAHRTARS 486
 DB 411 TDTGAVIEAIHEIDGTFTGLGLRYDRFKVKTHDGK-----TVSS 449

QY 487 FALSGNW--YFTPOHKLSTASHQ--ERLPSTQE-LYAHGKH---VATNTPEVGNKHLN 537
 DB 450 SSINFSFGVWQPREHWSFSASHNVASRFLYDALQTHKKGIIISADGT-----K 501

QY 538 KERSNNIELALGYEGDRWQYNLALYR-----NRFNGYIYAQTLNDG----- 578
 DB 502 ABRARNTGIFNVDGTFAANGSYFQTIKDALANPQRHDSVAVREAVNAGYIKNHGYE 561

QY 579 -----RGPKSTEDDSEMKIYRNSQGFYGAE-GEIY-----EKP 613
 DB 562 LGASYTGTGLTAKVGSRKPRFYTHPKLLSANPE-----FGAQTGRWTASLAYRFPK- 616

QY 614 TPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFIADQDNAPRVPAAARLGFHLKASLT 673
 DB 617 NPNLEIGWGRYV-----QKATGSLAAGQKDRGKLENVVRQGFVN--- 659

QY 674 DRIDANLDYRYPQAOKLARYETRTPGHMLNLGANYRNTYRGWVYKADNLLNSV 733
 DB 660 -----DVFANWKPIGKOTLN-----VNLSVN-----NVDFREY 687

QY 734 YAHSSFLSDT-POMGRSFTGGVNVKF 758
 DB 688 YPHSQRTWNTLPGVGRDVRIGVNYKF 713

RESULT 20
 ABP77241
 ID ABP77241
 XX standard; Protein; 713 AA.
 AC ABP77241;

XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 1012.
 XX KW Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB02069.
 XX PR 12-FEB-2001; 2001GB-0003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX DR WPI; 2003-058415/05.
 XX DR N-PSDB; ABZ38211.
 XX FT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX PT medicament for treating or preventing N. gonorrhoeae infection -
 XX PS Disclosure; Page 262; 815pp; English.
 XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
 XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
 XX CC antibodies that specifically bind to the proteins. The composition
 XX CC comprising the protein, nucleic acid or antibody is useful for the
 XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 XX CC infection, this may be in the form of a vaccine or gene therapy.
 XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 XX CC molecules of the invention.

QY 11 LSILLINTPL-LAQAHETEQSGLETVTVVVKSRPRATSGLLHTSTASDKLIISGDTLRQK 69
 DB 8 LSLLSLTAAAGFAHAENANVALDTVTVKGRQ-----GSKRTNIV---TLQOK 55

QY 70 ----AVNLGDALDGVPGIHASQYGGASAPVI-----RGOTGRRIKV-----LNHH 111
 DB 56 DESTATDMRELLKEEPSI---DFGGNGTSGFTLRGMGQNSVDIKVDNAYSQSILYHQ 112

QY 112 GETGDMADFPDHAIVDTALSOQVEILLRQPVTVLLYSSGNVAGLVADVADKIKEKPE-- 169
 DB 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDAQ 153

QY 170 ---NGVSGELGLRLSSGNLEKLTSGGINIG---LGK---NFVLHTEGLY---RKSGDYAV 217
 DB 154 DLLKGLDKXWVRLNSG---FAGNNGVSYGASVFGKEGNF---DGLFSYNRNDEKDYEA 206

QY 218 PR-YRNL---KRLPDSP-----RRFANGQHRVILGWRKPFYR---RTYSDRRQY 260
 DB 207 GKGFNRVNGGKTVPSYALDKRSYLAKIGTTFGDDHRIVLSHMKDQHRGIRTV---REEF 263

QY 261 GLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAH 320
 DB 264 AVGEN-----SRITIKRQAPAYRE---TTQSTNTNLAYTKDGLGFVEKLDANAY 309

QY 321 AHNGKPMIDLNRKRYELRAE---WKQPPGFPALRVHLNRNDYHDEKAGDAVENFNQ 377
 DB 310 V-----LEKKYSADDDKNGYAGNVKGNPNHTRATRCGMNFNDSRLAE-----Q 353

QY 378 T-QNARIELRHQ---PIGLKSGWGYOYLGQKSSAL-----SATSEAVKQPMLLDNKVQ 427

Query Match 5.4%; Score 216; DB 24; Length 713;
 Best Local Similarity 21.6%; Pred. No. 2.7e-10;
 Matches 187; Conservative 103; Mismatches 298; Indels 278; Gaps 47;

Db 354 TLLKYGINRHOEIKPQAFNSQFKIE---DKDATEEDKKNRNEKIAKAYRLTNPTK 410
 Qy 428 HYSFPGVQAN-WDNFTLEGGVVRKQKASIRYDKALIDRENYKQPLPDLGAHRQTARS 486
 Db 411 TDTGAYIEAIIHEIDGFTLTGLRYDRFKVKTGDK-----TVSS 449
 Qy 487 FALSGNW--YFTPOHKLISLTASHQ--ERLPSTOE-LYAHGKH-----VATNTEVGNKHLN 537
 Db 450 SSLNFSFGVWQPREHWSFSASHNVASRPDYDALQTHGKGKGIISADGT-----K 501
 Qy 538 KERSNNIELALGYEDRWQYNLALYR-----NRFNGYIYAQTINDG----- 578
 Db 502 AEARANTEIGFNYNDGTFAANGSYRQTIKDALANPQNEHDSVAVREAVNAGYIKNHGYE 561
 Qy 579 -----RGPKEIDDESEKLVRYNQSADFYGAE-GEIY-----FKP 613
 Db 562 LGASVYRTGGLTAKVGVSRSKPRFYDTHPKLLSANPE-----FGAQTGRWTASLAYREK- 616
 Qy 614 TPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAQPRVPAARLGFHLKASLT 673
 Db 617 NPNLEIGWRGRVY-----QKATGSILAAQKQKRDGKLENVVRQGFYV----- 659
 Qy 674 DRIDANLDYRVRFAQNKALRYETRTPGHHMLNMGANYRNRTRYGEWNNVVRADNLLNQSV 733
 Db 660 -----DVFAWKKPLGKDTLN-----VNLSVN-----NVFDKEY 687
 Qy 734 YAHSSFLSDT-PQMGRSFTGGVNVYKF 758
 Db 688 YPHSQRWNTTLFGVGRDVRVLGVNVYKF 713

RESULT 21

ID ABB78069 standard; Protein; 720 AA.
 AC ABB78069;

DT 05-NOV-2002 (first entry)
 DE Amino acid sequence of p64 polypeptide.
 XX p177; p88; p64; p55; p46; vaccine; gonorrhea.

KW Neisseria gonorrhea.
 OS WO200260936-A2.
 EN 08-AUG-2002.
 XX 31-JAN-2002; 2002WO-US02881.

PF 31-JAN-2001; 2001US-266070P.
 PR 06-AUG-2001; 2001US-310356P.
 PR 23-OCT-2001; 2001US-344452P.
 XX (IOWA) UNIV IOWA RES FOUND.
 PA (REGC) UNIV CALIFORNIA.
 PA (APIC) APICELLA M.A.
 PA (EDWA) EDWARDS J L.
 PA (GIBS) GIBSON B W.
 PA (SCHE) SCHEFFLER K.
 PA (BROW) BROWN E.

PI Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
 DR WPI; 2002-619227/66.
 DR N-PSDB; ABQ78300.

XX New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
 PT gonorrhea, useful for preventing, or protecting a female patient
 PT against, N. gonorrhea colonization or infection -
 XX

PS Claim 7; Page 117-120; 130pp; English.

CC The present sequence represents a p64 polypeptide. The specification
 CC describes p177, p88, p64, p55 and p46 polypeptides from Neisseria
 CC gonorrhea. The polypeptides are useful as vaccines, for preventing,
 CC or protecting a female patient against, N. gonorrhea colonization or
 CC infection. Such immunisation can prevent gonorrhea in women.
 XX

SQ Sequence 720 AA;

Query Match 5.3%; Score 214.5; DB 23; Length 720;
 Best Local Similarity 21.6%; Pred. No. 3.e-10;
 Matches 186; Conservative 101; Mismatches 315; Indels 259; Gaps 45;

Qy 16 INTPL-----LAQAHETEQSGLTETVTVGKSRPRATSGLLHTTASDKITS 62
 Db 1 MNTPLRLSLSLTLAAGFAHAENNAKVVLDTVTVKGDQ-----GSKIRTNIV- 50
 Qy 63 GDTLRQK-----AVNLGDALDGVPGIHASQYGGASAPVI-----RGQTGRRIKV----- 107
 Db 51 --TLOQKDESTATDMRELLKBEPSI---DFGCGNGTSQPLTLRGMGQNSVDIKVDNAYSD 105
 Qy 108 ---LNNHGETGDMADFSPDHALIWDLTALSOQVEILRGPVTLIYSSGNVAGLVDVADGKPT 164
 Db 106 SQILLHQGR-----FIVDPALVKVSVQVK-----AGSASAGIGATNGAII 146
 Qy 165 EKMPK-----NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVLHTGELY---RKSG 213
 Db 147 AKTVDAQDLLKGLDKNWGVRLNSGFASNEGSYGASVFGKEGNF-----DGLSYNRNDEK 202
 Qy 214 DYAVPR-YRNL---KRLPDSP-----RRFANGQRAVLGWRKFRFYR----- 251
 Db 203 DYEAGKGFNVNGKTVPEYSALDKRSYLAKIGTTFGDDDHRIVLSHMKDQHRGIRTVREE 262
 Qy 252 -TYSRRDQYGLPAHSHEYVDCHADIIWQKSLINKRYLQLYPHLLTEDEVDYDNLGSLCG 310
 Db 263 FTVGDKSRINIDROQAPAYR-----TTQSTNLTAYTGNLIG 299
 Qy 311 FHDDDDAHAAHNGKPIDLNRKRY---ELRAEMKQPPFGPEALRVHLNRNDYHDEKAG 367
 Db 300 FVEKLDANAYV-----LEKERYSADDSGTGYAGNVKGNHTRITRGANFNFDRLA 351
 Qy 368 DAVENFFNQOT-QNARIELRHQ---PIGRLLKSGWGVYQLGQSSALSATSEAVKQPM--- 420
 Db 352 E-----QTLKYGINRYRHOEIKPQAFNSKFSIPTTEKN-----GQKVDKPMEQQ 397
 Qy 421 -----LLDNKVQHYVSFFGVEQA--NWDNFTLEGGVVRKQKASIRYDKALIDR 466
 Db 398 MKDRADEDTVHAYKLSNPTKTDGTGVYVEAIIHIDIGDFTLTGLGRYDRFKVK--THDKTVSS 456
 Qy 467 ENYKQPLPDLGAHRQTARSFALSGNWYFTPOHKLISLTASHQERLPSTOEIYAHGKH--- 523
 Db 457 SNLN-----PSFGVIWQPE-----HWSFSASHNY---ASRSPRL---YDALQTHGKRGII 501
 Qy 524 -VATNTEVGNKHLNKRNSNNIELALGYEDRWQYNLALYRNRFGNYIYAQTINDGRGPK 582
 Db 502 SIADGT-----KABEARNTEIGFNYNDOGTFAAN-----GSYEW-QTIKDALANP 544
 Qy 583 SIEDDSEMKLVRYNQSADFYGAE-GEIYFKPTPRYRIGVSGDYVRGRLKNLPSLPGRED 641
 Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASFTYTGGLTAKVGVS-----HSKPRFYD 593
 Qy 642 AYGNRPFIADDDQNAQPRVPAARLGFHLKASLTDRI-DANLDY-YRVFAQNKALRYETRTPT 699
 Db 594 THDKLLSANPER-----GAQVGRWTASLAYRFPQNPLEIGWR-----GRYVQKAT 640
 Qy 700 GHEMLNLGANYRR-----NTRYGEW-----NMYVKADNLLNQSVYAHSS 738
 Db 641 G-SILAAQKQKRDGKLENVVRKGFVNDVFANWPKLGDITLNVLSVNNVFNKYYFPHSQ 699
 Qy 739 FLSDT-PQMGRSFTGGVNVYKF 758
 Db 700 RWNTLFGVGRDVRVLGVNVYKF 720

[illegible]

36 W. 110th St.

QY 352 RYHLNRNDYHDEKAGDAVENFFNNOTONAFIELRHQPIGRLKGSGWVQLGOKSSALSA 411
:
:
:
Db 358 KYGIN---YRHOEIKPQA---FLNSQ---FKIEQEKATDEEKNK-----NR 395
:
:
:
QY 412 TSEAVKQPMILDNKVQHYSFPGVEQAN-WDNFTLEGGRVVRVEKOKASIRVDKALIDRENY 470
:
:
:
Db 396 ENEKIAKAYRLNPTKTDTGVAYIEAHEIDGFTLTGLRYDRFKVK-THDGKTVSSNNLN 454
:
:
:
QY 471 KQPLPDLAGRAHTARSFALSNGWNVYFTPPOHKLSLTASHQEKLPSLTOELYAHGKH----VAT 526
:
:
:
Db 455 ----PSFGVIWQPE-----HWSPFSASHNY---ASRSPL--YDALQTHGKRGIISAD 499
:
:
:
QY 527 NTFEVGNKHLNKERSNNIELALGYEBGDQRKQYNLALYRNRFQNIYYAQTLNDG-RGPKSIE 585
:
:
:
Db 500 GT-----KAERARNTIEGFYNDGTFAAN-----GSVPW-QTIKDALANPNQRH 542
:
:
:
QY 586 DDSEMK-----LVRYNOSGADFYGARGEGIFYPKTPPYRI 619
:
:
:
Db 543 DSVAVREAENAGYIKNHGYELGASYRTGGLTAKVGVSHSKPRFYDTHDKLLSANPEFGA 602
:
:
:
QY 620 GVSQDVYRG---RLKNLPSP-----GR--EDAYGNRPFIADODDONAPRVPAARLGPHLKA 670
:
:
:
Db 603 QVGRWTASTLAYRFQN-FNLBIGWRGTRYVQKAVGSILVAGQDRNGKLENVVRKGFVN-- 660
:
:
:
QY 671 SLTDRIDANLDYRVYFAQNKUARVETRTPGHMLNLGANYERNTRYGEWNWYVKADNLLN 730
:
:
:
Db 661 -----DYFANWKPLCKDTLN-----VNLSVN-----NVFN 685
:
:
:
QY 731 QSYVAHSSFLSDT-PQMGRSFTGGVNVXF 758
:
:
:
Db 686 TFYPHPSQRWTNLTLPVGRDVLGVNYKF 714
:
:
:

RESULT 26
AAAY75568
ID AAY75568 standard; Protein; 703 AA.
XX
AC AAAY75568;
XX
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2610.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
XX WO9957280-A2.
XX
XX PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
FR 31-JUL-1998; 98US-0094869.
FR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX

(CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Morà M,
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.

DR	N-PSDB; AA254330.	Novel Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics -	Claim 2; Page 1238; 1453pp; English.
XX	AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941	represent novel <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> polynucleotides and polypeptides. AA254537 to AA254576 and AA254616 to AA254743 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to <i>Neisserial</i> bacteria (e.g. <i>meningitis</i> and <i>septicemia</i>), to detect the presence of <i>Neisseria</i> bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.	
XX	Sequence 703 AA;		
XX	Query Match 5.2%; Score 208.5; DB 21; Length 703;		
XX	Best Local Similarity 20.6%; Pred. No. 1.3e-09;		
XX	Matches 181; Conservative 119; Mismatches 259; Indels 321; Gaps 52;		
QY	12 SILLINTPLLA---QAHETEQSGLTETVTVGKSRPRATSGLLH-----TSTADKLIIG- 63		
DB	12 TLIIASFPAAADTQNGEHTATLTFSVVGQSDTSVLKGIYINDEAAVTRNGQLIKET 71		
QY	64 ---DTLR-QKAVNLG-----DALDGVPGITHASQYGGASAPVIRGOTGRRIKVLNHGGE 113		
DB	72 PQITIDTLNKKNNYGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117		
QY	114 TGMADFSPDHA-----IMVDTLASQVEILLRGPVTLTYSSGNVAGLVADVADGKIEKMP 168		
DB	118 QADASDIYRDGVRESQVRRSTANIERVELIKPFSVLVYGRRTGGGVINMVSKYANFKOS 177		
QY	169 EN-----GVSGLGKRLSS--GNLEKLTSG--GINIGLGNFVLHT 205		
DB	178 RNTGTVGSWANRSLNMDINVLNKNVAIRLTGEVGRANSFRSGIDSKNVVSPSITVKL 237		
QY	206 EGYLRKSGDYAVPRYRNLRKLPD--SPRRFANGQRAVLGWRKRFYRTYSDRDQVGLP- 263		
DB	238 DNLGLKWTGQYT--YDNVERTPDRSP-----TKSVY-----DRFGLPY 272		
QY	264 ---AHSHEYDDCHADIIWQKSLINKEYLQLYPHLLTEEDVDYDNPGLSGCFHDDDDAHA 319		
DB	273 RMGFARHNDP-----VKDKLQVW-----RSDLEY----- 296		
QY	320 HAHNGKFPWIDLRNRYELRAEWKQPPFGFBLRVHLNRNDYH---DEKAGDAVENFF-- 374		
DB	297 -AFNDK-W-----RAQWQ-----LAHRTAAQDFDHFYAGSENGNLIKRNAYW 336		
QY	375 ---NNQTONARIELR-HQPIGRKLG--SWGQVYLQK-----SSALSAT----- 412		
DB	337 QQTDNKTLSNLTNGDYTIIGRFENHLTVGMDSYRSREHNPFTLGFSSAFSASINPYDRASW 396		
QY	413 -SEAVKQPMLLDN--KVQHSYFF--GVEQANWD-NFTLEGGVRVEKQASIRYDVKALIDR 466		
DB	397 PASGRLOPILTONRHKADSIGIFVQNIFSATPDLKFKVLGG-----RYDKYTFNS 445		
QY	467 ENYKQPLPDLAGHQ--TARSPA--LSGNWYFTPOHKLSLTASHQRLPSTQBELYANGKH 523		
DB	446 ENKL-----TGSRSQVSGHSFSPNICAVWVNIPIVH--TLTASYNK--GFAPYGGRGY 494		
QY	524 VAINTEVGNKHLNKRSENNIELAL--GYBGDRWQYNLALYRNRFGNVYIAQTINDRGSP 581		
DB	495 LSIDTLSSAVFNADPEYTRYETGVKSSWLDRLSTLTLSAY-----YRIGVSGD----- 535		
QY	582 KSIEDDSEMLKRYNOSGADFYGAEBIYFKPTPR---YRIGVSGD----- 624		

```

Db 536 -----QIERFN-----IRYDPKNNPIYAYGKSRGVELSALGQI 574
Qy 625 -----YVRGLKNLPSLPGRDAYGNRPPIAQDQNAAPRVAARLGFHLKASLTDRIDAN 679
Db 575 IPKLYLRGSLGVWQAK-----VVEDKENPDRV-----GTHL--NNTSNVTGN 615
Qy 680 LDYR-----VFAQNKLYARET-----PGHMLN--LGANYR--NTRYGE 718
Db 616 L-FRYTPTTENLYGEIGVTGKRYGYSRKEVITLPGFARYDAMLGNHKNVNTFA- 673
Qy 719 WNWVVKADNLLNQSVYAHSSFLSDTPQMGSRFTGCVNVKF 758
Db 674 -----AANLFNQKYWRSDSM-----PGNPGYTARVNVRF 703

RESULT 27
AAB68916
ID AAB68916 standard; Protein; 703 AA.
AC AAB68916;
XX
XX 18-APR-2001 (first entry)
DE Neisseria meningitidis protein #15.
XX Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
KW rni5; rth; tolC.
XX Neisseria meningitidis.
OS
XX EP1069133-A1.
XX 17-JAN-2001.
XX 13-JUL-1999; 99EP-0401764.
XX 13-JUL-1999; 99EP-0401764.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX Nassif X, Tinsley C;
XX WPI; 2001-082916/10.
XX N-PSDB; AAF56456.
XX Immunogenic polypeptides derived from Neisseria meningitidis and the
XX nucleic acids that encode them, useful for diagnosing and vaccinating
XX against Neisseria infections e.g. bacteraemia and meningitis -
XX Claim 3; Fig 15B; 240pp; English.
XX
XX The present invention provides the protein and coding sequences of
XX several genes from Neisseria meningitidis. These include the dsaA, fhaB,
XX fhuA, rni5, rth17, rth18, rth19, rth20, rth21 and tolC genes. These can
XX be used in the diagnosis and treatment of infection by the bacterium,
XX which can lead to meningitis and bacteraemia, and in vaccines to prevent
XX such infection.
XX
XX Sequence 703 AA;
XX
Query Match 5.2%; Score 208.5; DB 22; Length 703;
Best Local Similarity 20.6%; Pred. No. 1.3e-09;
Matches 181; Conservative 119; Mismatches 259; Indels 321; Gaps 52;
Qy 12 SILLINTPLA---QAHETEQSVGLTIVTVKSRPRATSGLLH-----TSTASDKIIG- 63
Db 12 TLIIAFPPVAADTQNGEYHATLPTVSVGQSDTSVLKGYINDEAAVTENGQLIKET 71
Qy 64 ---DTLR-QKAVNLG-----DALDGVPGIHASQVGGGASAPVIRGOTGRRIKVLNHGE 113
Db 72 PQTIDTINQKNKYGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117

```

```

Qy 114 TGDMAFSPDHA-----IMVDTALSQQVEILRGPVTLTYSSGNVAGLVVDVADGKIPERP 168
Db 118 QADASDIYRDGVRESGVRRSTANIERVEILLKGPSSVLYGRNGGVINMWKYANFQOS 177
Qy 169 EN-----GVSGELGLRSS--GNLEKLTSG--GINIGLGNFVLHT 205
Db 178 RNIGTVYGSWANRSLNMDINEVLNKNVAIRLTGVEGRANSFRSGIDSKNVMVSPSITVKL 237
Qy 206 EGLYKSGDYAVPRVNLKRLPD--SPRRFANGQHRVILGWKRFRYRRTYSRRDOYGLP- 263
Db 238 DNGLKWTGGYT---YDNVERTEDRSP-----TKSVY-----DFGLPY 272
Qy 264 -----AHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTBEVDYDNPGLSCGHHDDDAHA 319
Db 273 RMGFAHRNDF-----VKDKLQVW-----RSDLEY----- 296
Qy 320 HAHNGKPIDLNRKRYELRAEWKQPPGFPEALRVHLNRNDVHH---DEKAGDAVENFF-- 374
Db 297 -AFNDK-W-----RAQWQ-----LAHTAAQDFDHFYAGSNGNLIKNYAW 336
Qy 375 ---NNOTONARIELR-HQPIGRKLG--SWGVOYLQK-----SSALSAT----- 412
Db 337 QQTDNKLSSNLTNGDYITIGRFENHLTVGMDYSREHNPFTLGFSSAFSASINPYDRASW 396
Qy 413 -SEAVKQPMLLDN--KVOHYSFF--GVEQANWD--NFTLEGVVRVEKQKASIRYKALIDR 466
Db 397 PASGRLOPILTQNRHKADSYGIFVQNIQSATPDLKFLVGG-----RYDKYTFNS 445
Qy 467 ENYVKQPLPDGAHRO--TARSFA--LSGNWYFTPOHKLSTASHOERLPSTQELVAHGKH 523
Db 446 ENKL-----TGSSROYSGHSFSPNIGAVNNIPVH--TLIASYNK--GFAPYGGRGY 494
Qy 524 VATNTEVGNKHLNKRNSNNIELAL--GYEGDRWQYNLALYNNRFGNYIYAQTLNDGRGP 581
Db 495 LSIDTLSSAVFNADPEYTRQYETGVKSSWLDRLSTLSAY----- 535
Qy 582 KSIEDDSEMKLVRYNQSGADFYGAEGEIVFKTPR---YRIGVSGD----- 624
Db 536 -----QIERFN-----IRYDPKNNPIYAYGKSRGVELSALGQI 574
Qy 625 -----YVRGLKNLPSLPGRDAYGNRPPIAQDQNAAPRVAARLGFHLKASLTDRIDAN 679
Db 575 IPKLYLRGSLGVWQAK-----VVEDKENPDRV-----GTHL--NNTSNVTGN 615
Qy 680 LDYR-----VFAQNKLYARET-----PGHMLN--LGANYR--NTRYGE 718
Db 616 L-FRYTPTTENLYGEIGVTGKRYGYSRKEVITLPGFARYDAMLGNHKNVNTFA- 673
Qy 719 WNWVVKADNLLNQSVYAHSSFLSDTPQMGSRFTGCVNVKF 758
Db 674 -----AANLFNQKYWRSDSM-----PGNPGYTARVNVRF 703

RESULT 28
AAB68923
ID AAB68923 standard; Protein; 702 AA.
XX
XX AAB68923;
XX
XX 18-APR-2001 (first entry)
XX
XX Neisseria meningitidis protein #22.
XX Meningococcus; meningitis; bacteraemia; vaccine; debA; fhaB; fhuA;
XX rni5; rth; tolC.
XX Neisseria meningitidis.
XX EP1069133-A1.
XX 17-JAN-2001.
XX 13-JUL-1999; 99EP-0401764.

```

QY	602	FYGAGEYIFKPTPR---	YRIGVSGD-----	YVGEELKXLPSP	633
		: :	: :	: :	
Db	541	-----IRYRDPDKNNPIYAV	SGKSRGVELSAIGQIIPK	LYLRSGLGVMQAK---	590
		: :	: :	: :	
QY	640	EDAYGNRPPIAQQONAPRVP	AAARLGFHLKASLTDRI	DANLDYRV-----	685
		: :	: :	: :	
Db	591	-----VVEDKENPDRV---	GIHL--NNTSVTGNL--	FFRYTPTENLYGEIGVTG	633
		: :	: :	: :	
QY	686	-----FAQNKALARYETRT	PGHHMLN--LGANYRR-N	TRYGWNWYVKADNLLNQSV	738
		: :	: :	: :	
Db	634	TKGRYGYNSRKEVTTLP	GFARVDAMLGNHKNVTFE	-----AANLLNOKYWRSDS	686
		: :	: :	: :	
QY	739	FLSDTPQMGSRSTGCVNVKF	758		
		: :	: :	: :	
Db	687	M-----PGNPRGYTARV	NYRF 702		
		: :	: :	: :	
RESULT 29					
AAB68919					
ID	AAB68919 standard; Protein; 703 AA.				
XX	AAB68919;				
DT	18-APR-2001 (first entry)				
XX	Neisseria meningitidis protein #18.				
DE					
XX					
KW	Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;				
KX	rni5; rth; toIC.				
XX	Neisseria meningitidis.				
OS					
XX					
PN	EP1069133-Al.				
XX					
PD	17-JAN-2001.				
XX					
PF	13-JUL-1999; 99EP-0401764.				
XX					
PR	13-JUL-1999; 99EP-0401764.				
XX					
PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.				
XX	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.				
PI	Nassif X, Tinsley C;				
XX					
DR	WPI; 2001-082916/10.				
XX					
DR	N-PSDB; AAF56459.				
XX					
PT	Immunogenic polypeptides derived from Neisseria meningitidis and the				
XX	nucleic acids that encode them, useful for diagnosing and vaccinating				
PT	against Neisseria infections e.g. bacteraemia and meningitis -				
XX					
PS	Claim 3; Fig 18B; 240pp; English.				
XX					
CC	The present invention provides the protein and coding sequences of				
CC	several genes from Neisseria meningitidis. These include the dsbA, fhaB,				
CC	fhuA, rni5, rth17, rth18, rth19, rth20, rth21 and toIC genes. These can				
CC	be used in the diagnosis and treatment of infection by the bacterium,				
CC	which can lead to meningitis and bacteraemia, and in vaccines to prevent				
XX	such infection.				
SQ	Sequence 703 AA;				


```

Db 72 POTIDTLNKNKNGVNDLSSILEGNAGDA-----AYDMRGES-----IFLRGF 117
Qy 114 TGMADSPDHA-----IMVDTALSOQVEILRGVPTLLYSNGVAGLVVDVADGKIPKMP 168
Db 118 QADASDIYRDGVRESGQVRRSTANIERVEILKGPSSVLYGRTNGGGVINMVSYANFKQS 177
Qy 169 EN--GVSG-----ELGLRSS--GNLEKLTSG--GINIGLGNFVLHT 205
Db 178 RNIGAVYGSWANRSLNMDINEVLNKNVAIRUTGEVGRANSFRSGIDSKNVWVSPSITVKL 237
Qy 206 ECLYKSGDYAVPYRNKLKLPD--SPRFANGQHRVGLGMRKREYRRTYSDRDOYGLP-- 263
Db 238 DNLKWTGQYT---YDNVERTPDRSP-----TKSVY-----DRFLPY 272
Qy 264 ---AHSHEYDDCHADIIWOKSL--INRKYLYL--PHLLTEEDVDYDNFGLSCGFHD 313
Db 273 RMGFAHRNDF-----VKOKLQVW-----RSDLEY----- 296
Qy 320 HAHNGKPWIDLRNKRKYELRAEWKQFPFGFEALRVHLNRNDYHH---DEKAGDAVENFF-- 374
Db 297 -AFNDX-W-----RAQWQ-----LAHRTAAQDFHFYAGSENGNLKRYAW 336
Qy 375 ---NNOTQVARIELR-HOPIGRKLG--SWGVOYLQK-----SSALSAT----- 412
Db 337 QOTDNKTLSSNLTLLNGDYITIGRFENHLTVGMDYGREHNRNPTLGFSSAFSASINPYDRASW 396
Qy 413 -SEAVKQPMLLDN--KVQHSYFF--GVEQANWD--NFTLEGGRVVEKQKASIRYDKALIDR 466
Db 397 PASGRLQPILTQNRHKADAYGIFVQNIIFSATPDLKFVLGG-----RYDKYTFNS 445
Qy 467 ENYKQPLDGAHQ--TARSA--LSGNVYFTPOHKLSTASHQERLPSTQELIYANGKH 523
Db 446 ENKL-----TGSSQYSGHSPFNIGAVMNINPVH--TLVASYNK--GFAPYCGRGY 494
Qy 524 VATNTEFVGNKHLNKRNSNIELAL--GYEGRWQYNLALVRNFRGNVYIAQTLDNGRGP 581
Db 495 LSINTSSAVFNADPEYTRQVETGVKSSWLDRLSTLSAVQIERFNIRY-----P 546
Qy 582 KSIEDSEMKLVRYNQSGADFYGAGETYPKTPRYRIGVSGDYVGRGLKNLPSLPQRED 641
Db 547 DEONDPYTVAVGKHRSRGVLSAIGQIIPKKL-----YLRGSLGVNQAK----- 591
Qy 642 AYNRPFTAQDONAPRYPAARLGFHLKASLTDRIDANLDYRV----- 685
Db 592 -----VVEOKENPDRV-----GIHL--NNTSNVTGNL--FFRYTPTENLYGEIGVTGTG 636
Qy 686 --FAQNKLARYETRTFGHHMLN--LGANYRRNTRYGEWNWYVKADNLNQSVYAHSSFLS 741
Db 637 KRYGYSNRNKEVTTLPGFARVDAMLGNHKK-----NVNITFAAANLLNQK--YWRSDAMP 689
Qy 742 DTPQGRSFTGGVNVKF 758
Db 690 GAP---RTYTARVNSYF 703

RESULT 30
ID AAB68921 standard; Protein; 702 AA.
XX AAB68921;
AC AAB68921;
XX
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis protein #20.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
KW rni5; rth; toIC.
XX
OS Neisseria meningitidis.
XX
PN EP1069133-A1.
XX

```

```

Pd 17-JAN-2001.
XX
Pf 13-JUL-1999; 99EP-0401764.
XX
Pr 13-JUL-1999; 99EP-0401764.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nassif X, Tinsley C;
XX
WP1; 2001-082916/10.
DR N-PSDB; AAF56461.
XX
PT Immunogenic polypeptides derived from Neisseria meningitidis and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
PT against Neisseria infections e.g. bacteraemia and meningitis -
XX
PS Claim 3; Fig 20B; 240pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC several genes from Neisseria meningitidis. These include the dsbA, fhaB,
CC fhuA, rni5, rth17, rth18, rth19, rth20, rth21 and toIC genes. These can
CC be used in the diagnosis and treatment of infection by the bacterium,
CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
CC such infection.
XX
SQ Sequence 702 AA;
Query Match 5.1%; Score 205; DB 22; Length 702;
Best Local Similarity 20.7%; Pred.No. 2.7e-09;
Matches 178; Conservative 114; Mismatches 287; Indels 280; Gaps 50;
Qy 12 SILLINTPLLA---QAHETEQSVGLTFTVVGKSPRATSGLLH-----TSTASDKIISG- 63
Db 12 TLIASFPPVAAADTDNGEHTATLPTVSVGSDTSVLKGVINVDAAVTRNGQLIKET 71
Qy 64 ----DTLRQKAVNLG-----DALDGVPGIHASQYGGGASAPVIRGTGRRIKVLNHHGET 114
Db 72 POTIDTLNKNKNGVNDLSSILEGNAGDA-----AYDMRGES-----IFLRGF 117
Qy 115 GDMADSPDHA-----IMVDTALSOQVEILRGVPTLLYSNGVAGLVVDVADGKIPKMP 169
Db 118 ADASDIYRDGVRESGQVRRSTANIERVEILKGPSSVLYGRTNGGGVINMVSYANFKQSR 177
Qy 170 N--GVSG-----ELGLRSS--GNLEKLTSG--GINIGLGNFVLHT 206
Db 178 NIGAVYGSWANRSLNMDINEVLNKNVAIRUTGEVGRANSFRSGIDSKNVWVSPSITVKLD 237
Qy 207 GLYKSGDYAVPYRNKLKLPD--SPRFANGQHRVGLGMRKREYRRTYSDRDOYGLP-- 263
Db 238 NGLKWTGQYT---YDNVERTPDRSP-----TKSVY-----DRFLPYR 272
Qy 264 ---AHSHEY--DDCHADIIWOKSL--INRKYLYL--PHLLTEEDVDYDNFGLSCGFHD 313
Db 273 RMGFAHRNDFVKOKLQ---VMSDLEVAFNQWRAQWLAHRTAAQDFD----- 318
Qy 314 DDDAHAAHNGKFWIDLRNKRKYELRAEWKQFPFGFEALRVHLNRNDYHHDEKAGDAVENF 373
Db 319 ---FYAGSENG-----NLIKRYN---AWOOTDNKTLSSNFTLN--GDY-----TIGRF 358
Qy 374 FNNQTNQVARIELRHQ--PIGRKSGWGVQYLGQKSSALSATSSEAVKQPMLLDN--KVQHSY 430
Db 359 ENHLTVGMDYSREHNRNPTLYGRSFTVPINPYDRASWPASGRL--QPILTQNRHKADSYG 416
Qy 431 FP--GVEQANWD--NFTLEGGRVVEKQKASIRYDKALIDRENYKQPLDGAHQ--TARS 486
Db 417 IFVQNIIFSATPDLKFVLGG-----RYDKYTFNSENKL-----TGNSRQYSGHS 459
Qy 487 FA--LSGNWVFTPOHKLSTASHQERLPSTQELIYANGKHVATNTEFVGNKHLNKRNSNI 544
Db 460 FSPNIGAVMNINPVH--TLVASYNK--GFAPYCGRGYLSIDTSSAVFNADPEYTRQY 514

```


WPI; 1996-485781/48.
N-PSDB; AAT44518.

Genes encoding H. influenzae HxuC and HxuD surface-expressed protein(s) - useful in the prepn. of vaccines for children against H. influenzae infection

Claim 4; Page 111-117; 188pp; English.

The HxuC protein (AAW01460) of Haemophilus influenzae type b (Hib) strain DL42 is a 78 kDa outer membrane protein involved in the utilisation of low levels of free haem. It shows homology to TonB-dependent outer membrane proteins of other bacteria, and was identified from an open reading frame (see also AAT44518) located upstream of the Hib DL42 hxa gene. Another outer membrane protein, HxuD (AAW01463), was also identified. Recombinant HxuC and HxuD proteins can be produced in transformed host cells and used to prepare vaccines for children against Hib infection; to raise diagnostic antibodies; and to prepare Hib diagnostic or therapeutic compns.

Query Match 5.0%; Score 203; DB 17; Length 725;
Best Local Similarity 19.8%; Pred. No. 4.4e-09;
Matches 159; Conservative 134; Mismatches 300; Indels 206; Gaps 43;

Qy 11 ISILLINTPLAQAHETEQSGLTETVVGKSRPRATSGLLHTSTASDKIISDGLTRKA 70
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 6 LSLAIATT--LVTANALAQSVELDSINVIATRDP---SRFATPEKQSK---DSLRSQA 57
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 71 VNTGDALDGVPGHASQYGGG---ASAPVIRQTGR--IKVLHHGTGDMADFPDH-- 124
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 58 TSVAAALEIDPNVDIR--GGSRIAQKPNIIGLSNRNVGVVIDVRQNFDLA-----HRG 110
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 125 AIMVDFTALSQOVELLRGPVTLVSSGNAGLVDAVGKIPKMPENGVSSELRLSSG- 183
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 111 SYFLPMSLIQEIEVIKGPSLSSGALGVGMARPTNALDLKKNN---DKFGVKIROGY 167
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 184 NLEKLTSGGINI-GLCKNFVLHTEGLRKSGDYAVPYRNLKRLPDSPRFANGQHRAV 241
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 168 QTANNLSERDVSPAANDKFVLLISGFYNMADNLTGKGKLN--TAVKQFGG---LAK 222
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 242 LGWRKRYRT-YSDRDQVGLPAHSHEYDDCHADIWKSLINKRYLOLYPHLLTEDV 300
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 223 FGWQINDANVELSHRETFKQAPS--NEVENELTNEQITDQIR 266
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 301 DYDNPGLSGCFHDDDDAHAHNGKPMWILRNKRYELRAEMKQPFGPEALVHLNRNDY 360
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 267 EFHPK-----NNGSP-----PKAPSQEP-----YGVKTRGVSYSY 299
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 361 HHDEKAGD-AVENFF--NNQTNARIELRHQIOGLKSGWGVOYLQOKSALSATSEA 415
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 300 LTDQIQIPQSTVFNYVLPDNPVNLTHIALYNNKT-----IEKBQKVSGVDQTKL 351
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 416 VKQPMILDN-KVQHYSF-FGVEQANWDNFTLEGV-----RVKQKA----- 455
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 352 TTRGINLSELSHISFVGVDFMR-DKIRTERGTNKDAQPADPYNANNTGVYLI 410
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 456 -----SIRD-----KALIDENRYKQLPDLGAHQRTARSFALSNWTF 495
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 411 AHIPLFGEKLLSPSYVDHYDTSSKTVKYKDNL-----SPATKLTIW 455
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 486 TFQHKLSLTASHOE--RLPSTOEIYANGKHVAT-----NTFFVGNHKNERSNII 545
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 456 T--NWLDPTAKYNAFAFPSQMERFVSGSHFGTSILGRNEINF-VANPNLARPETAKKE 512
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 546 L-----ALCYEDRWYNLALYRNFNGYVAOTLNDGRGPKSTDSDSEMKLVRNOS 598
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 513 ITANHFSLSLKQDGKFKIETAYFNVDKOFINLKIFNDAKTYNASASA-----CA 564
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 599 GADFYGAEGETFYKPT-PRYE-----IGVSGDYVRGL---KNLPSLPQREDAYGN 645
||| : | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

KW	antibacterial.
XX	
XX	Neisseria gonorrhoeae.
CS	Neisseria meningitidis.
OS	
XX	
XX	
PH	Key Location/Qualifiers
FT	32..36
FT	/label= TonB_box
XX	
XX	US6265567-B1.
XX	
XX	24-JUL-2001.
PD	
XX	
XX	05-APR-1996; 96US-0628434.
PF	
XX	
XX	07-APR-1995; 95US-0418964.
PR	
XY	(UYNC-) UNIV NORTH CAROLINA.
PA	
XX	
XX	
PI	Sparling PF, Beucher M;
XX	
XX	WPI; 2001-450739/48.
DR	N-PSDB; AAD10201.
DR	
XX	
XX	
PT	An isolated nucleic acid encoding a Fe-regulated protein B (FrpB),
PT	useful for production of a vaccine against infection by Neisseria
PT	gonorrhoeae or N. meningitidis -
XX	
XX	Claim 1; Column 25-30; 20pp; English.
PS	
XX	
XX	The present sequence is a Fe-regulated protein B (FrpB) isolated
CC	from Neisseria gonorrhoeae and Neisseria meningitidis (strain FA19). The
CC	present DNA sequence also represents the combined nucleotide sequence
CC	from PUNCH319 and PUNCH325. The FrpB is a 70 kD major iron-regulated,
CC	outer-membrane protein common to N. gonorrhoeae or N. meningitidis.
CC	The FrpB has particular use as a vaccine that protects a mammal from
CC	N. gonorrhoeae or N. meningitidis infection. The FrpB may also be used
CC	to detect the presence of antibodies specific for N. gonorrhoeae or
CC	N. meningitidis in a sample.
XX	
XX	Sequence 713 AA;
SQ	

[illegible]

AA68926	Standard; Protein; 703 AA.	QY	320	HAHNGKFWLDRNKRVELRAEWK-----QPPGFEALRVHLNRNDYHDEKAGDAVEN	372
AC	AA68926;	DB	297	-AFNDR-W-----RAQWLAHRTAAQDFHF-----YAGSENGNLKEN	333
XX		QY	373	FFNNQTQNAIIELR-----HQPIGRLLGK--SWGVOY-----LQOKSSALSATSEAVK	417
DT	18-APR-2001 (first entry)	DB	334	YAWQOTDNKTLSSNFTLNGDYITIGREFNHLTVGMDYSREHNPFTLG--SRAFTASIDPYD	392
XX	Neisseria meningitidis protein #25.	QY	418	-----OPMLDN--KVQHSFF--GVQANWD--NFTLEGGVVRKQKASIRYDKA	462
XX	Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;	DB	393	RASWPASGRLOPILTQNRHKADSYGFVQNIATPDLPKVLGG-----RYDK	441
XX	rni5; rth; tolC.	QY	463	LIDRENYIKOPLDGLAHQ--TARSPA--LSGNWYFTPOHKLSTASHOERLPSTOELYA	519
OS	Neisseria meningitidis.	DB	442	TFNSENKL-----TGSSRQYSGHSFSPNIGAVNNINPVH--TLYASYNK--GPAPYGG	490
XX	EP1069133-A1.	QY	520	HGKHVATNTEVGNKHLNKRSSNNIELAL--GVEGDRWQYNLALYRNRFGNYIYAQTLD	577
XX	17-JAN-2001.	DB	491	RGYLSINTSSSAVFVNADPEYTRQYETGVKSSWLDRLSTLSAYQIERFNRYR-----	545
XX	13-JUL-1999; 99EP-0401764.	QY	578	GRGPKSIEDDSEMKLVRYNOSGADFYGAEGEIVFKPTPRYRIGVSGDYVRGRLLNPLSLP	637
XX	13-JUL-1999; 99EP-0401764.	DB	546	---PDEQNDPYTWAVGKHSRGVLSAIGQIIPKXL-----YLRGSLGVWQAK--	591
(INRM) INSERM INST NAT SANTE & RECH MEDICALE.		QY	638	GREDAYGNRPFIADDDONAPRVPAARIGFHLKASLTDRIDANLDYRV-----	685
(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.		DB	592	-----VVEDKENPDV-----GIHL--NNTSNVTGNL--FFRYTPTENLYGEIGV	632
Nassif X, Tinsley C;		QY	686	-----PAQKLAHYETRTPGHMLN--LGANVRRNTRYGEMWYKADNLLNQSVAHS	737
WPI; 2001-082916/10.		DB	633	TGTCKRYGYSNRKKEVTILEGFARVDAMLGNHKK-----NVNITFAANLLNOK--YWRS	685
N-PSDB; AAF56466.		QY	738	SFLSDTQMGSRSPFTGGVNVKF 758	
Immunogenic polypeptides derived from Neisseria meningitidis and the		DB	686	DAMPGAP---RTVTARVNVSF 703	
nucleic acids that encode them, useful for diagnosing and vaccinating					
against Neisseria infections e.g. bacteraemia and meningitis -					
Claim 3; Fig 25B; 240pp; English.					
The present invention provides the protein and coding sequences of					
several genes from Neisseria meningitidis. These include the dsbA, fhaB,					
fhuA, rni5, rth1, rth18, rth19, rth20, rth21 and tolC genes. These can					
be used in the diagnosis and treatment of infection by the bacterium,					
which can lead to meningitis and bacteraemia, and in vaccines to prevent					
such infection.					
Sequence 703 AA;					
Query Match	5.0%; Score 200.5; DB 22; Length 703;				
Best Local Similarity	21.6%; Pred. No. 7e-09;				
Matches 186; Conservative 105; Mismatches 287; Indels 283; Gaps 50;					
QY	12	SILLINTPLLA---QAHETEQSVGLETVTVVWKSRRPRATSGLIH---TSTASDKLIISG-	63		
DB	12	TLIIASFVVAADTDNGEHYATLPTVSUVGSDTSVLKGYNDVDEAAVTRNGQLIKET	71		
QY	64	-----DTLR-QKAVNLG-----DALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHGGE	113		
DB	72	POTIDTLNIQKNKYGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF	117		
QY	114	TGMADFSPPHA-----IMVDTALSOQVEITLRGEVTLILYSSGNVAGLVADVADGKIPEKMP	168		
DB	118	QADSDIYRGVRESGGVRRSTANIERVELIKPSSVLGRTNGCGVIVNWKYANFKOS	177		
QY	169	EN--GVSG-----ELGLRLSS--GNLEKLTSG--GINIGLGRNFVLHT	205		
DB	178	RNIGAVTVGNARSRLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNVMVSPSITVKL	237		
QY	206	EGLYRKSGDYAVPRYNLKLDP--SPRRFANGQRAVLGWRKFFRYRTYSDRRDQGLP-	263		
DB	238	DNGUKWTGQYT---YDINVERTPDRSP-----TRSVY-----DRGLPY	272		
QY	264	-----AHSHYDDCHADIIWOKSLINKRYLQLYPHLLTEEDVDYDNFGLSCGFHDDDAHA	319		
DB	273	RMGFARHNDF-----VKDLQVW-----RSDLEY-----	296		

Grandi G, Masiugnani V, Pizza M, Rappuoli R, Scarlato V;
WPI; 1999-327407/27.
DR N-PSDB; AAZ12259.

PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection
 XX
 PS Claim 4; Page 379-380; 524pp; English.

XX Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*
 CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX Sequence 725 AA;
 Query Match 5.0%; Score 200; DB 20; Length 725;
 Best Local Similarity 21.7%; Pred. No. 8.2e-09;
 Matches 178; Conservative 105; Mismatches 294; Indels 244; Gaps 43;

QY 19 PLLAQAH-----ETQSVGLTETVTVVVGKSRPRATSGLLHTST-----AS 57
 DB 15 PVYQAADVSVDDPKQESDELTTITVADRTASSNDGYTVSGTHPLGLPMTLREIPQS 74
 QY 58 DKISGDTLR-QKAVNLGDALDGVPGHISQYG---GGASAPVIRGOTGRRIRKVLNHGCE 113
 DB 75 VSVITSQMDQNKIKTLDRALLQATGTSRQIYSGDRAGYNILFARGS-----KIANQIN 129
 QY 114 TGMADFSPPHAIMVDLTALSCQVEILRGPVTLTYSSGNVAGLVADVADGKIPKMPENGYS 173
 DB 130 GIPVADALADTG-NANTAAVERVEVVRGVAGLLDGTGEPATVNLVKRLTRK-PLFEVR 187
 QY 174 GELGLRLSSNLEKLTGGNIGLGNKFNVLTEGLRKSGDYAVPRYRNKLRPLDPSRPF 233
 DB 188 AEAGNRKHFG-LDADVSGSLN-----TEGTLR-----GRIVST---F 220
 QY 234 ANGOHRAVLGRKFRFRTYSDRRDQYGLPAHSHEVD-----DCHADIIWQKSLINKRYL 288
 DB 221 GRGD-----SNR-----RRERSDAELXGI---LEYDIAPTQTRVHAGMDYQQA----- 260
 QY 289 QLYPHLLTEBVD-----YDNLGLSCGFHDDDAHAHANGK-----PWIDLR-NK 333
 DB 261 -----KETADAPLSYAVYVDSQYATAFGPKDNPATNWNANRHRALNLFAGIEHRFNQ 312
 QY 334 RYELRAEW-----KOPFPGEALRVHLNRNDYHDEKAGDAVENFF---NNOTQVARI 384
 DB 313 DWKLKABEYDTRSPRQPYGVAGVLSI-----DHNTAATDLIPGVHADPRTHSASVS 365
 QY 385 L--RHQPIGR-----LKGSGVQVYLGQSSALSATSEAV-----KOPMLLDNKV 426
 DB 366 LIGKYLFGREHDLIAGINGKYKASNYGERSIIPNAIPNAYEFSRTGAYPQSPASQTI 425
 QY 427 QHY-----SFFGVEQANWNETLEGVVRVEKOKASIRYDKALIDRENYKQPLDGA 479
 DB 426 POYTRQIGYLAIRFADNLSLILGRTYRTG-SYDSR-----TGMYTVSA 476
 QY 480 HROTARS---FALSGNWFYTPQKLSLTASHQERLPSTQELYARHGKHVATNTEVGNKHL 536
 DB 477 NRFTPYTGIVFDLGN-----LSLYGVSYSFLFVPSQKDEHGSYLKPVIT----- 520
 QY 537 NKESNNIELALGYED-----RWQYNLALYRNRENYIYAQTLDGEGPKSIEDSEMKL 592
 DB 521 ----GNNLE--AGIKGEWLEGRNLASAAVVRARONNLATAA-----GRDP----- 559
 QY 593 VRYNQSADFY-----GAEGEYFKPTPRVEI---GVSGDYVR---GRKLNPLSLPG 638
 DB 560 ----SGNTYRAANQAKTHGEVEVGRITPEWQIQAGYSQSKTRDQSGRLNPDVSPE 614
 QY 639 RE----DAYGNRPPIAQDDQNAVRPAPARLGFHLKASLTDRIDANLDYRVFQNKIARY 694
 DB 615 RSFKLFTAY-----HFAPEAPS---GWTIGAGVRWQSETHDTPATIRIPNPAKA 661
 QY 695 ----ETRTPGHMLNLGANYRNTRYGEMWNYVKADNLNQ 731

DB 662 RAADNSROKAVAVADIMARYRFPNRA---ELSLNVDNLFNK 699

RESULT 39

AAE05851
 ID AAE05851 standard; Protein; 687 AA.
 AC AAE05851;
 DT 24-SEP-2001 (first entry)
 DE *Pseudomonas stutzeri* open reading frame-K (ORF-K) protein.

XX Open reading frame-K; ORF-K; pyridine-2,6-bis (thiocarboxylate);
 KW PDC; environmental remediation; phytoremediation; bioaccumulation;
 KW water purification; solution mining mobilisation; immobilisation;
 KW detoxification; redox state modifier; metal ion reactivity;
 KW carbon tetrachloride; metal.

XX *Pseudomonas stutzeri*.

OS WO200153309-A1.

PN 26-JUL-2001.

PD 19-JAN-2001; 2001WO-US02386.

PF 20-JAN-2000; 2000US-0177251.

PR (IDAH-) IDAHO RES FOUND INC.

PA (PASZ/) PASZCZYNSKI A.

PA (SEBA/) SEBAT J L.

XX Paszczynski A, Sebat JL, Lewis TA, Crawford RL, Cortese MS;

XX WPI; 2001-465361/50.

DR N-PSDB; AAD11171.

XX New isolated nucleic acids, useful for producing enzymes required to

PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for

PT reducing the amount of metal or carbon tetrachloride in a substrate,

PT e.g. soil or water

XX Claim 33; Page 95-98; 172pp; English.

XX The present sequence is a *Pseudomonas stutzeri* open reading frame-K
 CC (ORF-K) protein. The *Pseudomonas stutzeri* genome includes ORFs that
 CC encode enzymes required for biosynthesis of pyridine-2,6-bis
 CC (thiocarboxylate) (PDC). The ORFs encoding PDC are especially useful
 CC in environmental remediation methods, e.g. phytoremediation,
 CC bioaccumulation, water purification, waste water purification, solution
 CC mining mobilisation, immobilisation, detoxification, redox state
 CC modifier or modification of metal ion reactivity. In particular, the
 CC ORFs are useful for degrading carbon tetrachloride and removing metals
 CC from substrates, e.g. soil or water.

XX Sequence 687 AA;

Query Match 4.9%; Score 198.5; DB 22; Length 687;

Best Local Similarity 21.4%; Pred. No. 1e-08;

Matches 170; Conservative 111; Mismatches 286; Indels 229; Gaps 43;

QY 7 KPVLISILLINT---PLLAQ---HTEQSVGLTETVTVVVGKSRPRATSGLLHTSTASD 58

DB 21 QPMWATALICAFVPGPQLAFAPGSAASPDSTTLPEITVTAETKTERP---LERVPASV 76

QY 59 KIISGDTLRQKAVNLGDALDGVPGHISQYG---GGASAPVIRGOTGRRIRKVLNHGCE 116

DB 77 AVIDCWADEQSGITSLKQLEGRIPLGSPFQGMNSFVNRGL-----TAN 123

QY 117 MADFSPDHAIWVD---TALSQ-----QVETLRGPVTLTYSSGNVAGVADVADGKI 163

Db 124 PNFSSSTLLIADGVPTLTQAFBSGMLDLRIEIVIRGPSTLYGRNAEAGV--IAHSL 181
 Qy 164 P-EKMPENGVSGLGLR-----LSSGNLEKLTSGGI--NIGLGNFVLHTEGLYRKS 212
 Db 182 PMDATPRASVAEAGSRNKRVMRFALSOPLVEERLYGVSNGWSSQDGFIDNTH-TGKA 240
 Qy 213 GDYAVPRYRNKRLPDSPPRFANGCHRAVLGWRKRFYRTYSDRRDQYGLPAHSHEYDDC 272
 Db 241 DD-----REQNL-NLGLRPAAGATDVV--NRYAHQYDDGASLWSPGAPRK-RVA 289
 Qy 273 HADIWQSLNKRILQYLPHLLTEEDVDYDNPGLSCGFHDDDAHAHANRKPWIDLNRN 332
 Db 290 SGTSPSWNRS-----EGQTLSPNFVQHFASGLRLHSVTAWNEFKD 328
 Qy 333 KYELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNQOTQARIELRHQIGR 392
 Db 329 -----RIQDQTFMPADVL--HVGED--HH-----LRTLSEFRVE-----GQ 362
 Qy 393 L-KGSW--GVQVLGQKSSALSATSEAVKQPMLLDNKVQHYFFGVQANWD-----NFTL 444
 Db 363 LGEASWLAGV-YADRSNDLHSTSTKTMWG--LSDIRADQDQSDTAALFTHNVPILSADWSI 419
 Qy 445 EGGVVEVEKASIRYDKALIDRENYYKQPLDGLGAHROTARSFALSGWNYFTPQHKLSLT 504
 Db 420 DAGARVERNEVLQ-RQATSEKQWTHVSPRLALHQ-----IT 458
 Qy 505 ASHQBELPSTOELYAHGKHVATNTPEVGNKHLNKNERSNNIELALGYEG-----DRWQYNLA 560
 Db 459 ANHQWYVASASRVRTGGFNLAFTL---GYLPVDTKNWSYETGLKGLWLLDKIRYSLA 514
 Qy 561 LYRNFNGYVIAQTLDNDRGPKSIEDSEMKL---VRYNOSG--ADPYGAGEIYFKPT 614
 Db 515 AY-----LMD-----IDMQVQMPVGVWMIITSAATATSKGLELDV----- 551
 Qy 615 PYRIGVSGDYVRG---BLKNLPS-----LPREDAYGNRPFFIAQD-----DON 655
 Db 552 -----DYLLGGGWQLKGLAWNHTRFDFHRDGEADYDGNQNPFPAPDLTGHLGIRYD 602
 Qy 656 APRVPAARLGFHLKSLDRIDANLDYRVFAQNKLAARYETETPCHHMLNLGANYRNTR 715
 Db 603 APE-----GWAQASVTSKSSKYLDAANGYERN-----GYGLNVLVAGYQ--- 643
 Qy 716 YGEMNNYVKADNLLNQ 731
 Db 644 -GNWEIAAVADNATDQ 658
 RESULT 40
 ID AAY34501
 AC AAY34501;
 AC AAY34501;
 XX
 XX 20-MAR-2003 (updated)
 DT 25-AUG-1999 (first entry)
 DE
 DE Porphyromonas gingivalis protein PG40.
 XX
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphyromonas gingivalis.
 XX
 XX W09929870-A1.
 XX
 XX 17-JUN-1999.
 XX
 XX 10-DEC-1998; 98WO-AU01023.
 XX
 XX 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 PR 30-JUL-1998; 98AU-0004963.
 PR 04-AUG-1998; 98AU-0005028.
 XX (CSLC-) CSL LTD.
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX WPI; 1999-385613/32.
 DR N-PSDB; AAX91719.
 DR Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX
 PS Claim 1; Page 482-483; 588pp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 708 AA;
 Query Match 4.9%; Score 198.5; DB 20; Length 708;
 Best Local Similarity 20.1%; Pred. No. 1.1e-08;
 Matches 163; Conservative 103; Mismatches 302; Indels 241; Gaps 35;
 Qy 15 LINTPLLAQAHETEQSV-----GLETVTVVGSKRPRATSGLLHTSTASDKIISGDTL 66
 Db 8 LLSIGISAQAFKTDNVTDSLRVHNLQTVTVYS-----TRTAVPLKKIPARKMELISSRNI 63
 Qy 67 RQKAV-NLGDALDGVPIHASOYGGASAPVIRG--QTGRRIKVLNHHGETGDMADFSFD 123
 Db 64 KQSGFNWMTDLTKQSSLDVIQYFSSNIGIRGFKPSKYTVLVNGIPAG-----TD 117
 Qy 124 HALMVTALSQQVEILRGVPTLLYSSGNVAGLVVDVADGKIPEKMPEN-----GVSG 174
 Db 118 NISTLNTSNIEQIEILKGPFSIYGTNMGVVNIITHKSKDKIHGNVSLFGGSYQTMAG 177
 Qy 175 ELGLRLSSGNLEKLTSGGINIGLKG-----NF-----VLHTEGLYRKS 212
 Db 178 SFNL---GGRFEDIPFSDISLGLDKQNDKYKTSNNFLSKLEEAIVDVNATKKNKMG 234
 Qy 213 GDYAVPRYRNKRLPDSPPRFANGCHRAVLG-----WRKRFYRTYSDRRDQYGLP 263
 Db 235 SDYVATGRLRFGIDFTPEWSNLNYQNVLGDAIPVGGSIW-----GVYGESKKLNRS 288
 Qy 264 AHSHEY---DDCHADIIWQSLNKLKEYLQYLPHLLTEEDVDYDNRN--PGLSCGFHDD---- 314
 Db 289 STSFELLGKHGCHT-----LQFSPYFNTEKSENYNADPTGFIYKSDYYTY 335
 Qy 315 -----DDAHAAHNGKFWIDLNRN-----KRYELRAEWKQFP-PGF----- 348
 Db 336 GALLQDKISFGGQIVLGVDSRNMTWESERFEQAGVNTKPYNPGVATNIGLFGQANFYL 395
 Qy 349 --EALRVHLN-RNDY--HHDEKAGDAVENFNQOTQARIELRHQIPGRUKSGWQVYLQ 404
 Db 396 LNDALISAGARADFMFFDLKANEXYLNNEAKQETHNV-----INPNVGIKYEFV 444
 Qy 405 KSSALSAT--SEAVKQPMLLDNKVQHYFFGVQANWD-----NFTLEGGRVREKOKASIR 458
 Db 445 KGLTAGHTFGSAPSAFQKAGQYVGPFGTTIGNFDLPKPEKSMTWDFGIGYSNARCGIO 504

QY 459 YDKALIDRENYVKQPLDGLGAHRQTARSFALSGNWYFTPOHKLSTASHOERLPSTQELY 518
 Db 505 ADVTLT-----YFTHDKOLILSSPD-----Y 526
 QY 519 AHGKHVAINTFEVGNKHLNKNERSNNIELALGYEGDRWOYNLALYRNRFNGYIYAQ-TLND 577
 Db 527 ANNITTYINA-----DKARMSGIEALLSYDFG-----SLFANKFSLRAFANATIML 572
 QY 578 GRGPKSIEDD---SEMKLVRYNQ--SGADFYGARGEIYFKPTPRYRIGVSGDYVRGLKN 632
 Db 573 NSEMKKSOTDAPWSEMYVRKONITFGIEYRKEG-----LEVMLNGRFMGRRIE- 622
 QY 633 LPFLPGRDAYGNRPFFIAQDDQ-----NAPRVPA-----ARLGFHLKASLTD 674
 Db 623 -----QNWYAYYVEVRPELQOLLAABEPELAAQGLLEHPQAMVFNASAYYHMKYLT 675
 QY 675 RIDANLDYRVFAQNKLARYETRTFGHHM 703
 Db 676 GVNLN-----NILDELYTEKDGYHM 695

Search completed: November 14, 2003, 10:58:40
 Job time : 55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:56:57 ; Search time 21 Seconds

(without alignments)

3471.231 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLSILLINTPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3845	95.3	764	2 D81883	probable outer-mem
2	694.5	17.2	681	2 E82812	outer membrane hem
3	679	16.8	687	2 D83547	hypothetical prote
4	636	15.8	687	2 E87275	TonB-dependent rec
5	290	7.2	713	2 F82506	probable TonB syst
6	264.5	6.6	715	2 H90977	hypothetical prote
7	264.5	6.6	719	2 F85824	hypothetical prote
8	246	6.1	721	2 E83237	probable TonB-depe
9	243	6.0	680	2 H70347	outer membrane pro
10	231.5	5.7	764	2 H83055	probable outer mem
11	231	5.7	652	2 D82317	iron-regulated out
12	228	5.6	665	2 AD0160	probable outer mem
13	228	5.6	744	2 B64049	outer membrane pro
14	227	5.6	652	2 S25265	outer membrane pro
15	224	5.6	723	2 F83173	outer membrane pro
16	221	5.5	723	2 C40508	outer membrane pro
17	214.5	5.3	720	2 S81833	Fe-regulated prote
18	213.5	5.3	714	2 F81962	probable iron-regu
19	212.5	5.3	703	2 D81215	TonB-dependent rec
20	212.5	5.3	714	2 C81018	iron-regulated out
21	211	5.2	673	2 A56148	pepticin receptor
22	211	5.2	673	2 AF0232	pepticin/yeastiniab
23	210	5.2	673	2 S60142	outer membrane pro
24	208.5	5.2	703	2 E81792	probable TonB-depe
25	208	5.2	710	2 E83360	conserved hypothet
26	205.5	5.1	676	2 T12069	hemin receptor - y
27	205.5	5.1	676	2 AG0035	hemin receptor pre
28	203	5.0	725	2 A57148	outer membrane pro
29	203	5.0	802	2 AC2075	outer membrane hem

30	202	5.0	713	2 A56268	Fe-regulated prote
31	200.5	5.0	751	2 AC0574	ferrienterobactin
32	200	5.0	725	2 H81030	TonB-dependent rec
33	199	4.9	616	2 E83485	probable TonB-depe
34	192	4.8	725	2 D81976	probable ferric si
35	191	4.7	625	2 AD0476	probable vitamin B
36	188	4.7	660	2 D91176	heme utilization/t
37	188	4.7	660	2 E86022	outer membrane hem
38	187.5	4.6	599	2 AC3334	metal chelate oute
39	187.5	4.6	663	1 QREIC	colicin I receptor
40	187.5	4.6	690	2 A80124	probable TonB-depe
41	185	4.6	614	2 A88241	hypothetical prote
42	185	4.6	614	2 F86088	hypothetical prote
43	184.5	4.6	659	2 G91009	colicin I receptor
44	182.5	4.5	614	2 AB0935	vitamin B12 recept
45	181	4.5	686	2 S28042	hemin receptor pre

ALIGNMENTS

RESULT 1

D81883
probable outer-membrane receptor protein NMA1161 [imported] - Neisseria meningitidis (s
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: D81883
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: D81883
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-764 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CA884423.1; PID:G73798
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA1161

Query Match	95.3%;	Score 3845;	DB 2;	Length 764;
Best Local Similarity	95.9%;	Pred. No. 6.2e-266;		
Matches 730;	Conservative 7;	Mismatches 18;	Indels 6;	Gaps 3;
Qy	1	MAQTTLKPIVLSILLINTPLAQAHTEQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI	60	
Db	7	MAQTTLKPIVLSILLINTPLAQAHTEQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI	66	
Qy	61	ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIRKVLNHHGETGMDAP	120	
Db	67	ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIRKVLNHHGETGMDAP	126	
Qy	121	SPDHAIMVDTALSOQVEILLRGPVTLTYSSGNVAGLVVDVADGKIPEKMPENGVSGLGLRL	180	
Db	127	SPDHAIMVDSALSOQVEILLRGPVTLTYSSGNVAGLVVDVADGKIPEKMPENGVSGLGLRL	186	
Qy	181	SSGNLEKLTSGGNIIGKQNVLTGKSGDYAVPRYRNKLPLDSPFRFANGQHPA	240	
Db	187	SSGNLEKLTSGGNIIGKQNVLTGKSGDYAVPRYRNKLPLDSPFRFANGQHPA	243	
Qy	241	V-LGW--RRFRVRRYTSRRDQVGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE	297	
Db	244	IGLSWVGKGFIAAYSRRDQVGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE	303	
Qy	298	EDVDYDNPGLSCGFHDDDAHAHANGKWDILRNKRYELRAEWKQPPFGPEALRVHLNR	357	
Db	304	EDIDYDNPGLSCGFHDDDAHAHANGKWDILRNKRYELRAEWKQPPFGPEALRVHLNR	363	
Qy	358	NDYHDEKAGDAVENFNNTQWARIELRHQPIGRKLSGWQVYLGQKSSALSATSEAVK	417	
Db	364	NDYHDEKAGDAVENFNNTQWARIELRHQPIGRKLSGWQVYLGQKSSALSATSEAVK	423	

```

QY 418 QPMLLDNKKVHYSTFGVEQANWNTLEGGVYRVEKQASIRYDKALIDRENYKQPLPDL 477
DB 424 QPMLLDNKKVHYSTFGVEQANWNTLEGGVYRVEKQASIRYDKALIDRENYKQPLPDL 483
QY 478 GAHROTARFALSGNWNFFYPQHKLSLTASHQBELPSTOLYAHGKHVATNTFEVGNKHLN 537
DB 484 GAHROTARFALSGNWNFFYPQHKLSLTASHQBELPSTOLYAHGKHVATNTFEVGNKHLN 543
QY 538 KERSNNIELALGEGURWQYNLALYRNRFGNIIYAQTLDNGRGPKEIDDDSEMKLVRYNQ 597
DB 544 KERSNNIELALGEGURWQYNLALYRNRFGNIIYAQTLDNGRGPKEIDDDSEMKLVRYNQ 603
QY 598 SGADFVGAEGEIVKFTPRYRIGVSDYVGRGLKNLPSLPGREDAYGNRPFTAQDQNP 657
DB 604 SGADFVGAEGEIVKFTPRYRIGVSDYVGRGLKNLPSLPGREDAYGNRPFTAQDQNP 663
QY 658 RVPAAELGPHLKASLTDRIDANLDYRVPFAONKLARYETRTPGHMLNLGANYRNTRYG 717
DB 664 RVPAAELGPHLKASLTDRIDANLDYRVPFAONKLARYETRTPGHMLNLGANYRNTRYG 723
QY 718 ENNNYVKAADNLNQSVYAHSSFLSDTPQMRSGFTGGVNVKP 758
DB 724 ENNNYVKAADNLNQSVYAHSSFLSDTPQMRSGFTGGVNVKP 764

RESULT 2
E82812
Outer membrane hemin receptor XF0384 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 23-Dec-2002
C:Accession: E82812
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82812
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-681 <SIM>
A:Cross-references: GB:AE003890; GB:AE003849; NID:g9105215; PIDN:AAF83194.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canargo, L.E.A.; Carraro, D.M.; Carrer, H.
as-Neto, E.; Docena, C.; El-Dofry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0384
C:Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to
Query Match 17.2%; Score 694.5; DB 2; Length 681;
Best Local Similarity 27.4%; Pred. No. 2.3e-41;
Matches 220; Conservative 115; Mismatches 294; Indels 173; Gaps 22;
QY 1 MAQTTLPVLSILLINTPLLAQA-----HETQSVGLEVTYVVG----- 40
DB 9 LPRVFLVAVSSLL--TPALAMADVCPPTDSHFXTTEVTHLKDLDAVVVTAIPLRD 66
QY 41 ----KSRPRATSGLLHTSTASDKTIGSDTLRQ-KAVNLGDALDGVFGIHASQYGGGASAP 95
DB 67 SYSDLRSFVA-----LLAGERLDEVRASSLIGETVAVLPGVQSSNPGVGRP 113
QY 96 VIRGQTGRIRKIVNLHGHGTGMDPSPDHALMVDLTALSQQVEILRGVPTVILYSSGNVAGL 155

```

```

DB 114 IIRGLDPRVAVLNLNGSLSSQDSTVSQDSHPAVEPFLANQIEVLKGPSILLYSGSAGGV 173
QY 156 VDVAADKIPKMPENGVSGBELGLRLSSGNLEKUTS--GGINIGLGKNFVLHTEGLYKSGD 214
DB 174 VNIVDGRIAE-APVGGFNGRAEMRLDGDGKHGNTNMFRIADAGNGSALSADGVYREKX 232
QY 215 YAVPRYENLAKLPDPSRRFANGQHRVAVLWKRKFYRTYSDDRDQYGLPAHSHEYDDCHA 274
DB 233 YDFPKGQVNSPFDIT-----KSGSMGASFGSGDFGFLGSVARFHDYSG----- 275
QY 275 DIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNAHKPWIDLENKR 334
DB 276 -----NPG-----EPGDPVAGDRGSMRLRHQDR 298
QY 335 YELBAEMKQPPGFEALRVHLNRNDYHDEKAGDAVENFENNQTONARIELRHQPTGLRK 394
DB 299 YDLKAGLTDPWGBASALRFLSGHTQYDHIHIEFEGNEVGTTFGKRASEGRVEASFAFGGWR 358
QY 395 GSWGV-----QYLQKSSALSATSEAVKQPMLLDNKVQHYFFGVGEQANWNTLEGG 447
DB 359 TAFVQGGSDSTFQALGEESFVPKTNKSI-----GVFGLAHTNTFGLFQAEPG 405
QY 448 VRVEKQKASIRYDKALIDRENYKQPLDLCARHQTARSPALSGNWNFFYPQHKLSLTASH 507
DB 406 ARGDQ-----VKYDTDNGVTRNYH-----PGSLAFSGDLALSQWELTLNVDH 448
QY 508 QERLPSTOEIVAHGKHVATNTFEVGNKHLNKRSSNNIELALGYEGDRQWYNLALYRNRFG 567
DB 449 AERAPVEEELFAKQPHIATLAYEVRADLKKERANQALGLVFNENDSDAKVSTYYSRYG 508
QY 568 NYIYAQTLDNGRGPKEIDSEMKLV--RYNQSADPYGAEGEIVF-----KPTPRYRIGV 621
DB 509 NFIIY---LVD--TGTWTFWDDQDRDLPVRQWSQANAI FHIIEGEATFHLAKNTSGSMWLRV 564
QY 622 SGDVRGLKNLPSLPGREDAYGNRPFAODDQNAVRVPAARLGFHLKASLTDRIDA--- 678
DB 565 PGDTVSGRLKN-----GMLPRIVPARYGAELRWE-----DAGWR 599
QY 679 -NLDYRVRFAONKLARYETRTPGHMLNLGANYRNTRYGEMWYVYKADNLNQSVYAH 737
DB 600 TSLSAKRYEKQKNVAVNETTAGYTMVDLAHLAYHIDVDSIAWEVFFDGNLTLNRDARVHT 659
QY 738 SFLSDTPQM-CRSFTGGVNVKP 758
DB 660 SFLKDDVMLAGRNTYTAGLRMPF 681

RESULT 3
DB3547
Hypothetical protein PA0781 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: DB3547
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; E
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lir
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: DB3547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE004513; GB:AB004091; NID:g9946669; PIDN:AA04170.1; GSPDB:GN00
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0781

Query Match 16.8%; Score 679; DB 2; Length 687;
Best Local Similarity 27.4%; Pred. No. 2.9e-40;
Matches 212; Conservative 121; Mismatches 319; Indels 122; Gaps 25;

```

Qy 6 LKPIVLSILLINTPLL-AQAHETEQSGLTETVTVVKGSRPRATSGLLHTSTASDKIISGO 64
 Db 15 LTPALACLVSGEITLGDGPSELPSQVITANPLGNESPATPS-----SVLEGD 64
 Qy 65 ---TLQKAVNIGDALDGVPGHISGAGAPVIRGOTGRRIKVLNHHGTGMDAPSP 122
 Db 65 ELTLRQKG-SLGETLNGLPVGSSTVFGPGASRPVIRGMDGRIILNRNGVGLDASSLSY 123
 Qy 123 DHAIWVDTALSOQVEILLRGPVTLVSSGNVAGLVADVADCKIPEKMPENGVSGLGLRLSS 182
 Db 124 DHAVPEDNSVERLELVGPAALYGGNAIGGVNMSFDRIPSE-PVDGIHSGELRYGG 182
 Qy 183 GNLEKLTGGINIGLKPNVLTHTGLYKSGDYAVPRYRNKRLPDSPPR-ANGQHRA 240
 Db 183 ADTTRSRSALAEAGDGF-NFALHVDAAAREFNDVIRPGVAHSSR-----QRQIDGDTGKHV 237
 Qy 241 VLGWKRFRYRTYSDRRDQYGLPAHSHEVDDCHADI IWQSLINKRYLYQLYPHLLTDEDV 300
 Db 238 ---QNSGRQDGGVAGVSYHWHGYAGLSY-----SGYDS 269
 Qy 301 DYDNFGLSCGFHDDDDAHAAHNGKRWIDLRNKRVELRAEWKQPPFGFEALRVHLNRNDY 360
 Db 270 NYGSPA-----EDDVR-----LKWQODRYAFASEIRDLBGPFTSLKLDAAATKY 313
 Qy 361 HHDEKAGDAVENFNNOTQNAIRIELRHQPIGRKLSGWGVYLGQSSALSATSEAVKQPM 420
 Db 314 EHKEIEDGETGTTFKNGVEGRLEARHRLPLGLNGVGAQFANSRFSALG-----EEAF 367
 Qy 421 LLENKVKQHYSPFGVQANW---DNFTLEGGVVERKQASIRYDKALIDRENYKQPLD 477
 Db 368 VPHETETDSAALEEE-WKLSRDLQSGARLE---HTRVDPDAKGNRFAEND---417
 Qy 478 GAHRQTARSAFALSGNWFYTPQHKLSLTASHQBELPSTQELYAHGKHVATNTEVGNKHLN 537
 Db 418 GSCSFTTGLSGTGVAVYKLTPIWLSAATLSYTERAPTVELYANGPHAATGTVEVGDAAD 477
 Qy 538 KERSNNIELALGYEGDRWQYNIALYRNRFNGYIYAQTINDRGPK-----SIEDDSBMK 591
 Db 478 KEKAVSTDALRFDNGVHKSGVGVYRSPNVI---GLLASGRHNEEGEVVAAGDDEALP 535
 Qy 592 LVRYNOSGADFYGAEGE---TYFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRP 647
 Db 536 EYLYSGVADFYGVGAQDRHILLESYPYGNFDELSDGYTRAKNKD-----580
 Qy 648 FIAQDDQONAPRYPAARLGFHLKASLTDRIDANLDYRVFAQNKIARYETRTPGHMLNIG 707
 Db 581 ---TGEPLPRIAPLRDNTALIWEL-QQWQARVDVVEHAASQHRVPEEBELSTDGY---TTUG 633
 Qy 708 ANYRNRTRYGENW---YVKADNLLQSVYVAHSSFLSD-TPQMGSRFTGGVNVYKF 758
 Db 634 ASLGYNFDLGEGRWLAFAVKGTLNLTQTVRYASSILRDRVPAAGRGIEAGVKVAF 687

RESULT 4

E87275
 TonB-dependent receptor [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: E87275
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 n, J.; Laub, M.R.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: AB7249; MUID:21173698; PMID:11259647
 A:Accession: E87275
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-687 <STO>
 A:Cross-references: GB:AE005673; NID:gl13421339; PIDN:AAK22201.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0214

Query Match 15.8%; Score 636; DB 2; Length 687;
 Best Local Similarity 26.5%; Pred. No. 3.4e-37;
 Matches 206; Conservative 107; Mismatches 297; Indels 168; Gaps 23;
 Qy 22 AQAHETEQSGLTETVTVVKGSRPRATSGLLHTSTASDKIISGDTLRQKA-VNLGDALDGV 80
 Db 37 ASAQDAEVSKVITTAAPYGV-----ADALTASVAVLDRTDLOLAAPKGLGDALAGL 88
 Qy 81 PGHASOYGGASAPVIRGOTGRRIKVLNHHGTGMDADFSPOHAIMVDTALSSQVEILR 140
 Db 89 PGVRSITTFGAGASRPVVRGLAGPRVQVLTNGVQIDASALSPOHQTATDGEAERIEVL 148
 Qy 141 GPTVLYSSGNVAGLVADVADCKIPEKMPENGVSGLGLRLSSGNLEKLSGGINIGLKN 200
 Db 149 GPAALYGGSAIGGVNMIIDRISTQOPIIDGMSGRLLASRGTDGDSYALSAGVHATVGP- 207
 Qy 201 FVLHTEGLYKSGDYAVPRYRNKRLPDSPPRFA-----NGOHR-----AV 241
 Db 208 MVLTDALKRESKDYKIPY-----PESARQALAEGETAEGGAGRLSENSAVDLETFGAG 261
 Qy 242 LGM---RKRFRYRTYSDRRDQYGLPAHSHEVDDCHADI IWQSLINKRYLYQLYPHLLTEED 299
 Db 262 LSYVVDKGFVMSIKRTDSTYGVPGHAHE-----290
 Qy 300 VDYDNFGLSCGFHDDDDAHAAHNGKRWIDLRNKRVELRAEWKQPPFGFEALRVHLNRND 359
 Db 291 -----HEGEAEAGHBEESAVTIGLKQTRIDLRGEYDADLGPFPAKVPFSGHAD 338
 Qy 360 YHDEKAGDAVENFNNOTQNAIRIELRHQPIGRKLSGWGVYLGQSSALSATSEAVKQPM 419
 Db 339 YTHTEPEGDAVGTGFTSDGVEGRLELVQTERGCKWKGAVGVQALRRNRPDAIGDEAYVVK-- 396
 Qy 420 MLDNKKVQHYSPFGVQANW---DNFTLEGGVVERKQASIRYDKALIDRENYKQPLDGLA 479
 Db 397 ---TKITEGCAFTQORLELDGYEGGLRIDTRELDSLKGKA---DFTN-----439
 Qy 480 HRQTARSAFALSGN---WYFTPQHKLSLTASHQBELPSTQELYAHGKHVATNTEVGNKHX 535
 Db 440 -----LSGSVGAFWRPFTTSPFVGLALSRSRAPTESELFAGGPHAATRGFEIGDAD 490
 Qy 536 LNKERSNNIELALGYEGDRWQYNIALYRNRFNGYIYAQTINDRGPKSIBDDSSMKLVRY 595
 Db 491 LKBTATSLKATLHYGGERVSGDLHLIYARYDGFIDLRPTGD-----BEDG-LAVVRY 542
 Qy 596 NOSGADFYGAEGEITYFKPTPRYRIGVSG-----DYVRGLKNLPSLPGREDAYGN 645
 Db 543 VQDAEPRGFEEAEL-----AYRVMTDQSRVNLHAGADPVRG-----579
 Qy 646 RPTIAQDDQONAP-RVPAARLGFHLKASLTDRIDANLDYRVFAQNKIARYETRTPGHMLN 703
 Db 580 -----SSDLGPPARIIPPVSVS--LKASYQAPMWSGDVEVRRTGCGQERVAQWELPTDGYTV 632
 Qy 704 LALCANYR---RNTRYGENWYVVKADNLLQSVYVAHSSFLSD-TPQMGSRFTGGVNVYKF 758
 Db 633 LNAATLAWKPAGNARV---RULFDGRNLTNEEVREHVSFLKDIAPSPQGRVAGIALRF 687

RESULT 5

F82506
 Probable TonB system receptor VCA0064 [imported] - *Vibrio cholerae* (strain N16961 seroC
 C:Species: *Vibrio cholerae*
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: F82506
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
 J., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: AB2035; MUID:20406833; PMID:10952301
 A:Accession: F82506
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-713 <HEI>

RESULT 7
F85824

1

Db 528 -YIDFDELQVVS-NDVGNWNLGA-----TKHQIETSGHYDFAALDPRLDGSLVYG 577
 Qy 629 RLK-----NLPSLPGRDAYGNRFIAQDDQNAAPVPAABLGHKLKASLTDRIANL 680
 Db 578 SLTYTRATYEGDIFSGRDLPLYSR-----QVATAGVRYEV-----DRWTYNL 621
 Qy 681 DYRVFAQNKLYATETPTGHHMLNLGANYERN-----TRYGE-----WNVVYKAD 726
 Db 622 D---AFAQSM-----QAPAGLSTDSOG-NFTHNYITEPSADGQYGDIPGYVTNARVGYD 672
 Qy 727 -----NLINQSVYAHSS 738
 Db 673 GPQASNLKGLGVKNLFDKQYFTRSS 699
 RESULT 9
 H70347
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666; PMID:9537320
 A:Accession: H70347
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-680 <AQF>
 A:Cross-references: GB:AE000694; NID:g2983162; PIDN:AA06764.1; PID:g2983163; GB:AE000694
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: oprC
 A:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
 F:45-219/Domain: tonB-dependent receptor amino-terminal homology <TNN>
 Query Match 6.0%; Score 243; DB 2; Length 680;
 Best Local Similarity 21.0%; Pred. No. 3.5e-09;
 Matches 165; Conservative 117; Mismatches 326; Indels 178; Gaps 33;
 Qy 29 QSVGLETVVVVGKSRPRATSGLLHTSTASDKIISGDTLQK-KAVNLGDALD-GVPGHSHAS 86
 Db 17 QEVNLEEIQVIGKR-----EVLTEDVVRPELPKADVGAEVNIPIGVWKV 60
 Qy 87 QYGGGASAPVIRGQTRRIKVLNHHGETGDMADFPSPDHAIM-VDPALSQQVEILRGPVTL 145
 Db 61 RKGAINDVVRIGFKKDEVNQLFDGARVYNACPNRMDPGIEFHVDPSEVESVEVIKGPFD- 119
 Qy 146 LYSSGNVAGLVADVADKIPKMPKPVGSGELGRLSSGNLEKLTSGGINIGLGNFVLHT 205
 Db 120 VRYNAGVGTAVN-----KTEPKSGMEGRITTYADNWS-----TINPSYFSGKDRUSFL 171
 Qy 206 EG-LYKSGSDYAVPRYNLRLPDSPPRANQ-----HRAVLGWRKRYRRTYS 254
 Db 172 IGVAFFGKPYEDGKGRKITEIYSGNPSAYSQDERDSTAFNIHTA---WAKLRYKIDG 228
 Qy 255 DR-RQYQGLPANSHEYDDCHADILINQKSLINKRYLQLYPHLLTE---EDVDYDNPGLSCG 310
 Db 229 VRFKLDY---AHQRATD-----VLYPYLMDGIYDEVDRVNLGLE-- 265
 Qy 311 FHDDDDAHAAHNGKFWIDLRNKRVELRAEWKQPFPGFEALRVHLNRNDYHDEKAGD-- 368
 Db 266 -GKRFPKALYGVSVRHWMTNPKRVISQNAAPRGVTGCTVAKSVYGFKEG---SFGDFS 320
 Qy 369 -AVENFNNQTNARIELRHQIPIGRLKSGWGVQYLGQKSSALSTSEAVKQPMLLDNKVQ 427
 Db 321 FGIDTTFRYWKAQTMTYMNMGMYR-----TQNTIPDVVY 356
 Qy 428 HVSFFGVEQANWD-NFTLGGVVRVEKQKASIRYDKALIDRENNYKQPLPDLGAHQETARS 486
 Db 357 NFGLYGEYRKLSPKLRVLVAGLRDWSKTADSGKANINLYRY-----HNTDTTS 407

Qy 487 ---FALSGN---WYFTPOHKLSLTA---SHQERLPSTOELY-----AHGKHVATNTFEVGNKH 535
 Db 408 QTDVTPSGNVQVYFELKEGIELFAGLSAVRVPDQERFALDRMGKMENTMGDWGNPK 467
 Qy 536 LNKERSNRIELALGVGDRWQVNLALYRNRFQNYIYAQTLANDRGPKSIEDDSEMKLVRY 595
 Db 468 LDPERNNELDLGILKLTGERYSGLRTFLSYVKDYIYVVRATAPSSGNMWNQNT--KAMSY 525
 Qy 596 NOSGADFYGAEEIYFKTPRYRIGVSDYVRGLKNLPSLPGRDAYGNRPFFIAQDDQN 655
 Db 526 TNIDAYFYGFELSGSYAITDTIFFDGNVAYTRGR-----KDDTYPAKNITDKDIAE 576
 Qy 656 APRVPAARLGFH-----LKASLTDRIANLDYRVFAQNKLYATETPTGHHM 703
 Db 577 TPPL-TARLAARYDTGMYGQVEGVLAATQDNVSDLO-----EETSGYGI 622
 Qy 704 LML--GANYRNTRYGEWNVYKA--DNLLNOSVYAHSSFLSD-----TPQMGSRSTG 752
 Db 623 INLKGVNVR-----NLRVAGINNLFDKLYYTHLSYLRNPFRTGVKVPFGRTYYL 674
 Qy 753 GNVKVF 758
 Db 675 SVSYTF 680
 RESULT 10
 H83055
 A:Title: Probable outer membrane hemin receptor PA4710 [imported] - Pseudomonas aeruginosa (str. C); Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Dec-2002
 C:Accession: H83055
 R:Stover, C.K.; Pham, X.Q.; Brwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Li
 ; Lory, S.; Olson, M.V.
 A:Residues: 406-959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: H83055
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-764 <STO>
 A:Cross-references: GB:AE004485; GB:AE004091; NID:g9950968; PIDN:AAG08096.1; GSPDB:GNO
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4710
 C:Superfamily: iron-regulated outer membrane protein
 Query Match 5.7%; Score 231.5; DB 2; Length 764;
 Best Local Similarity 19.8%; Pred. No. 2.8e-08;
 Matches 172; Conservative 126; Mismatches 325; Indels 247; Gaps 35;
 Qy 6 LKPIVLSILIN-----TPLLAQAHETQSGVLETVTVVVGKSRPRATSGLLHT 53
 Db 9 LRP-CLALLLSPLSALAGNAVPLTPTTITATTEQAVDSVPSTVSVQTRQLD----- 61
 Qy 54 STASDKIISGDTLQKAVNLGDALDGVPGHSHASQVGGASAPV---IRGQTRRIKVLN 109
 Db 62 -----RQVNNIKELVRYEPGV--SVGGORAGITGVNIRKIDGNRLITQI 106
 Qy 110 HHGETGD---MADFPDHAIMVDTALSSQVEILRGPVTLIYSSGNVAGLVD----- 157
 Db 107 DGVLPNDPFGSPYQTHRYNVDPIVKKRVEILRGPSALYGSNAITGGAIVSYFTLDPDSI 166
 Qy 158 VADGKIPKMPKPVGSGELGRLSSG---NLEKLT-----CGINIGLGNFV 202
 Db 167 IKDGG-----DVCARLKAGYESASHSWLTSATVACRADDFDGLLHYGVYRQHE 214
 Qy 203 LHTEGLYRKSGDYAVPRYNLRLPDSPPRANQHRVILGWR-----KAFYR 251
 Db 215 TESNGGHHGTTG-----LSKSEANPEDADSYLLGLGNVYAGSRFGLVFEKYSD 265
 Qy 252 TYSDDRDQYGLPAHSHEYDDCHADI-----IWQK---SLINKRYLQLYPHLLTEE 298

```
Db      266 VDTDKSAYGGP-----YDKGKPAIPSPMLPGMYQMRKGNLTLTREYRGLHHFILDQS 320
Qy      299 DVYDNPGLSCGHDDDDAHAAHNGKPMIDLNRKRYELRAEWKQPPFGFEALRVHLNRN 358
Db      321 VADRIQWLSNLQAKTDQATREFY-----YPIIRKVLRLTR 355
Qy      359 DYVHDEKA---GDAVENFNNTONA---RIELRHQPIGLKSGWQVILGQKSSALSA 411
Db      356 DTYIKERLWVFDQOLDKSPAIGTEHLLSYGINKHOKVTGMRSGTGNLDTGADSPRDA 415
Qy      412 TSAVKQPMLLDNKVQHYSPFGVEQANWDFLLEGGVRVE-----KOKA 455
Db      416 LERSDDFP---DPTVKYALFAQDSISWNTWTFPGLRYDYTRMEPHITDEFLTKQSQ 472
Qy      456 STRYKALIDRENYKQPLDLAGHRTARSFALSGNWYFTPOHKLSLTASHQERLPSTQ 515
Db      473 NTAVDDES---DKMKHRVSPKFG---VTYDFAQHYTWY---GQYAOGF-----RTPAK 516
Qy      516 ELYAHGKHVATNTFEVGNKHLNKNERSNNIELALGYEDRWQYNLALYRNFNGYIYAQTL 575
Db      517 ALYGRFENLOAGHYEPNPNLKEKQSQFETGLRGKFDGSGFVAVFYNYRDFIDEDAL 576
Qy      576 N-----DGRGPKSIDDSMKLVRYNOSGADF-----YCGARGEIYFKPTPRYRIGVSG 623
Db      577 NTDSTGCGQTFCQSNRIERAVIK---GVELKGRLELGAFGAPQGLYTQGSVAYAYGRNK 632
Qy      624 DYVRGR-LKULPSLP-----GREDAVGNRPFIAQDDONAPRVPAARLGHFLKASLT---D 674
Db      633 D--NGEPIINSVNPFLTGVFGYDEADGN-----YGGLLSWTLVKRKD 672
Qy      675 RIDANLDYRVFAQNKRLARYETRTPGHHMLNGLANYR-----RNRTRYGEWN 720
Db      673 RVDDSS-----TFHTPDGTASQFTPGFGLVLDLSAYVRLSKDLTLNAGLYNLTKDKYWLMD 727
Qy      721 WYVKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db      728 DVRGYDSVGEASALAPAN-IDRLSQGRNF 756

RESULT 11
D82317
iron-regulated outer membrane virulence protein, TonB receptor family VC0475 [imported]
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82317
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82317
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 <HEI>
A:Cross-references: GB:AE004134; GB:AE003852; NID:g9654900; PIDN:AAF93648.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0475
A:Map position: 1
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog

Query Match 5.7%; Score 231; DB 2; Length 652;
Best Local Similarity 21.6%; Pred. No. 2.4e-08;
Matches 176; Conservative 108; Mismatches 310; Indels 222; Gaps 40;

Qy      1 MAQTTLKPIVLST---LLNTPLLAQ-AHETEOSVGLTWTVVVGKSRPRATSGLLHTSTA 56
Db      1 MSRENPSPVSLVTLGLMFSASAFADATKTDETM---VVTAGVYAQ-----VIQNAFA 51
Qy      57 SKKISGDTLRQKAV-NLGDALDGVPGIHASQVGGASAPV-IRG-----Q 100
```

```
Db      52 SISVISREDLESRYRDVTDALKSVPGVTVT---GGGDTTDISIRMGSENVTLLILVDGKRQ 109
Qy      101 TGRRIKVLNHHGTGDMADFPDPAHMDVTALSQOVELRGPVTLTLLSSGNVAGLVADVAD 160
Db      110 TSQTRP-NSDGPQIEQWLPPLQAI-----ERIEVIRGPMSTLYGSDAIGGVINIIT 161
Qy      161 GKIPKMPENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVPRY 220
Db      162 RKDQOQW-----SGNVQLSTVVOEN-----RASGD----- 186
Qy      221 RNLKRLPDSRRFANGQHRVAVLGRKRFYRTYSDRRDOYGLPAHSHEYDDCHADIIWQK 280
Db      187 -----EQSANFFVTGPLSDALS--LQVYQQTQRDEDEI-----EHGYGDKSL-----R 228
Qy      281 SLINKRYLQLYP--HLLTEEDVD---YDNPGLS-----CGFHDDDDAAHAH---AH 322
Db      229 SLTSKLYQLNPDHQLQLEAGVSAQDRENNVGVSAQSSGCGRTCSNTDQNYRRNHVAVSH 288
Qy      323 NGKFWIDLNRKRYELRAEWKQPPFGFEALRVHLNRNDYH--HDEKAGDAVENFNNQTON 380
Db      289 QG-----DWQD-----VQSDTYLYQYBENTNKSREMSIDNTVFK 322
Qy      381 ARTIELRHQPIGLKSGWQVILGQKSSALSATSEAVKQPMLLDNKVQHYGFFGVEQANWD 440
Db      323 STLV---APIGEHMLFGVE--GKHESLEDKTSNKISSRTHISN-TQWAGFIEDEWALAE 376
Qy      441 NFTLEGGVRVEKQASIRYDKALIDRENYKQPLDLAGHRTARSFALSGNWYFTPOHK 500
Db      377 QFRLTTFGRLDHDK-----NY-----GSH-----FSPRVYGVNMLDPLWT 411
Qy      501 LSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLNKNERSNNIELALGYE-GDRWQYNL 559
Db      412 VKGGVSTGRAPQRLRVTPDWGQVGGGNIYGNPDLKPKETSIKLSLMTSTGSLAASL 471
Qy      560 ALYRNRFP-----GNVYIAQTLNDGRGPKSIEDDSMKLVRYNOSGADFVGASGEI 609
Db      472 TAFHNDFKDKITVACPANICTAGPNQWGAAP-----TYRVNIDBAETYGABATL 521
Qy      610 YFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAAPRVPAARLGFHLK 669
Db      522 SLPITESVELSSSYTYTHSEQKS-----GNFAGRPLL-----QLPKHLFNANLS 565
Qy      670 ASLTDRID--ANLDY--YRVFAQNKRLARYETRTPGHHMLNGLANYRNTRYGEMWNVYKA 725
Db      566 WQTTDLRLNWANLNYRGKMQEGGASNDDFIAPSYTFIDTGVTYALTD-----TATIKA 620
Qy      726 D--NLLNQSV-YAHSSFLSDTPQMGRSFTGGVNVKF 758
Db      621 AVYNLFQEVNVAEYGVVED---GRYWLGLDIAP 652

RESULT 12
AD0160
probable outer membrane protein YPO1313 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD0160
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0160
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-665 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90143.1; PID:g15979363; GSPDB:GN00175
C:Genetics:
A:Gene: YPO1313
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homo

Query Match 5.6%; Score 228; DB 2; Length 665;
```

Best Local Similarity 20.7%; Pred. No. 4e-08;
Matches 152; Conservative 116; Mismatches 280; Indels 188; Gaps 34;
QY 24 AHETQSYGLTETVTVVGRSPRATSGL---LHTSTASDKIISGTLQKAV-NIGDLDG 79
Db 29 AEKNTATPTDMVY-----TASGFOQRIDSAASISVVTRQIENKAYRDVTDALD 81
QY 80 VPGIHASQYGGASAPV-IRGQTGRIRKIVLHGHGTGMADFP-----DHAIMVDTA 131
Db 82 VPGVVT---GGSTSDISIRGMAKYTLIL-VNGKRVTRTRPNSDGSQTEQWLPLA 138
QY 132 LSQQVEILRGPTVLLSYSGNVAGLVADVADGKIPEK-----MPENGVSGBGLRLS 181
Db 139 AIDRIEVRGPMSSLYGSDAMGCVINIIRKVGKEHGTVRADATLQBDSSGDI----- 193
QY 182 SGNLEKLTSGGINIGKFNVLHTGLY-RKSGDYAVPRYRNKLPDPSRPFANGQRA 240
Db 194 -FQTNAYASGLIDGL---LGLKVSGLLSHRSEDKIIDGY-----NQORMRNGTATP 241
QY 241 VLGWKRFYRRTYSRRDQYGLPFAHSHEYDCHADIIWQSLINKRYLQLYPHLLTEEDV 300
Db 242 TL-----TPDDNNEFDPDIGHYVDNRSTPGRTTLALNG---TNSDT 279
QY 301 DYDNFGLSCGTHDDDAHAHANGKFWIDLRNRYELRAEWKQPPFGEALRVHLNRDY 360
Db 280 QYDR-----NNYAVTHNG---YYDFGNS-----TSYIQRDET 308
QY 361 HHDERAGDAVENFFNQTNARIELRHOPIGRLKSGWGVYLGQKSSALSATSSEAVKQPM 420
Db 309 RNPQRKMSVDNIFN--TQTSFLDNHTLI--LGGQYRYEELYDKGNQLPSASDL----- 359
QY 421 LDNKKVQHSYFSGVEQANW---DNFTLEGVRYVEKOKASIRYDKALIDRENYKQPLD 477
Db 360 ---KKLTRWSWALFAEDQMTDNFALTGIRMDQD-----QNY----- 395
QY 478 GAHROTARGFALSGNWFYFPOKHLSTASHQRLPSTQ-----ELYAHGKHVATNTE 530
Db 396 GTH-WTPR---LYGVHLADQWLTKGVSQGYRSPDLQATDDWQLSGGKG-GLPALI 450
QY 531 VGNKHLNKRNSNIELALGYE-DRWQYNLALYRNFNYI-----YAQTLD 577
Db 451 LGNSNLKPSRSQEIIGILMDQEGNMASVTLYTDFDKDITEVRNCDITNTTQCCVF 510
QY 578 GRGPKSIEDSEMKLVYRNSQADFYGARGIEYFPTPRYRIGVSGDYVGRGLKNLPSLP 637
Db 511 GINYKFISS-----RINVDKAMTGAETAFANDINQAWSLATWYTFQSEKS----- 558
QY 638 GREDAYGNRPFTIAQDDQNAAPRVPARLGFHLKASLTDRIDANLDYRVFAQNKLAYPE-- 695
Db 559 ---GAPAGQPL---NQMPKHLNGTLNKTTEDEFATWIRAN---YRGKASBYLNRTSMG 608
QY 696 TRTPGHMLNLGANYR 711
Db 609 SRTPSYTFVDLGANYQ 624

RESULT 13
B64049
outer membrane protein hxC homolog - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64049
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: B64049
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-744 <TIGR>

A;Cross-references: GB:U32696; GB:L42023; MID:g1573057; PIDN:AAC21789.1; PID:g1573065;
Query Match 5.6%; Score 228; DB 2: Length 744;
Best Local Similarity 19.9%; Pred. No. 4.8e-08;
Matches 173; Conservative 125; Mismatches 318; Indels 252; Gaps 42;
QY 7 KPIVLSILLINTPLLAQAHEHQSGVLETVTVVVKSRPRATSGLLHTSTASDKIISGDTL 66
Db 7 KRIFLVLTVGLQVNTAKQSSNSRELLPI-----IVNTDSSNKLPGRSVL 56
QY 67 RQK-----AVNLGDALDGVPGIHASQYGGASAPVIRGOTGERIKVLAHGHGTGMADFP 120
Db 57 KQKNIEQXQADNANILNLPVGNMA---GGFRP---GGGT-----LNING-MGDAEDV 103
QY 121 S-----PDAIMVDTALSQQVEILRGPTVLLSYSGNVAGLVADVADGKIPEKM 167
Db 104 RVQLDGNATSKPEKYQOGSIFIEPELLRRTVDKGNYSPOYGVNGGFPAGTVKFEETKARDFL 163
QY 168 PEN-GVSGEIGLRLSSGNLEKLTSGGINI-GLGKNFVLHTEGLYKSGDYAVPRYRNKLR 225
Db 164 QENQKIGGFLKYGNNSNNQKTYSTALVLQNEQKNIDLLFGSVRNAGDY-----KR 215
QY 226 LPDSRPFANGQRAVL-----GWRKRFYRRTYSRRDQYGL 262
Db 216 PDNSKILFSGNOKTGLIKLNQISPEHLLTSSVYGIHKGW-----BPPAKRILPK 269
QY 263 PAHSH--EYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSGCFHDDDDAHAAH 320
Db 270 PSLSDIMRYG---TDIAWKKLVYR-----DQKDBNYTL-----KYN 304
QY 321 AHNGKWDILNRKRYELRAEWKQPPFGEALRVHLNRNDVHHDEKAGDAVENFFN----- 375
Db 305 LPENNPNINLSTQ-----FYSKT---TQNDMPKREASGLVSGLGNOSWIT 348
QY 376 -----NOTONARIELR-HOPIGRLKSGWGVYLGQKSSALSATSEAVK----- 417
Db 349 YSDLTPTDINTSTFNKTTVHELL-----FGLQWLKNTNTLMYDKSKVRKADYNYGF 402
QY 418 QPMLLDNKVQH-YSPFVEQANWNTFLEGVN-----VEKOKASIRYDKALIDRENY 470
Db 403 QPYTMPSGRQYQAFYLODQIKWKNIIFSTGVRYDHINNIQKNLAKYN----- 452
QY 471 KQPLDPLGALHQVARSFALSGNMY-----FTPOKHLSTASHQE---RLPSTQELYA---HG 521
Db 453 -----DISAGHDYSQKNYNGWSYLLGNYDVNHYLSLFTNFSKTWRAPVIDEYEQFKQ 507
QY 522 KHVAINTFFVGNKHLNKRNSNIELALGY---EGDRWQYNLALYRNFNYI----- 571
Db 508 SSVFATSLNLEKEMINQTRYGGI-ITLNLHLPQENDAFQFETTYFYNRGKNEIFKTRGVNC 566
QY 572 ---AQTLDNGRGPKEISDDSEMKLVRYNQSGADFYGARGIEYFPTPRYRIG-VSGDYVR 627
Db 567 VGNADNTNNKVCPLIEN-----YRNLPGYVYIQGALEAYYQST---YLFGEITYSYVK 617
QY 628 GRLKNLPSLP-GRBDAYGNRPFTIAQDDQNAAPRVPARLGFHLKASLTDRIDANLDYRVF 686
Db 618 GKRDTSRPNWGTSTW-----IA---EIPPRKATTALGENVPKY-----YLTVG 659
QY 687 AQNKLAVERTEP-----GHEMLNLGANYRNTRYGEW-----NMYVKADNLNQ 731
Db 660 WRABFVRQRQSPSLGDPKASSWSLPSRGSYLSHLNLFWSWPAKIKGMNVKITVDNLNFR 719
QY 732 SVYAHSSFLSDTPQMGSRFTG-GVNVKF 758
Db 720 AY-----NPVIGELASGTGRNKF 738
RESULT 14
S25265
outer membrane protein irga precursor - Vibrio cholerae
N;Alternate names: ferrienterochelin receptor homolog
C;Species: Vibrio cholerae
C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 08-Oct-1999

C;Accession: S25265; A37834
R;Goldberg, M.B.; Boyko, S.A.; Butterton, J.R.; Stoeber, J.A.; Payne, S.M.; Calderwood,
Mol. Microbiol. 6, 2407-2418, 1992
A>Title: Characterization of a Vibrio cholerae virulence factor homologous to the family
A:Reference number: S25265; UID:93023868; PMID:1406279
A:Accession: S25265
A:Molecule type: DNA
A:Residues: 1-652 <COL>
A:CROSS-references: GB:U72152; EMBL:M63192; NID:g1763224; PIDN:AAC44766.1; PID:g1763226
A>Note: The sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 299-1nch
R;Goldberg, M.B.; Boyko, S.A.; Calderwood, S.B.
J. Bacteriol. 172, 6863-6870, 1990
A>Title: Transcriptional regulation by iron of a Vibrio cholerae virulence gene and home
A:Reference number: A37834; UID:91072235; PMID:2174861
A:Accession: A37834
A:Molecule type: DNA
A:Residues: 1-152, 'D' <GO2>
A:CROSS-references: GB:M37773
C;Genetics:
A:Gene: irgA
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo
C:Keywords: membrane protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-652/Product: outer membrane protein irgA status predicted <MAT>
F:68-214/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:367-652/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

	Query Match	5.6%; Score 227; DB 2; Length 652;
	Best Local Similarity	21.6%; Pred. No. 4.6e-08;
	Matches 176; Conservative	108; Mismatches 310; Indels 222; Gaps 40;
Qy	1 MAQTILKPIVLSI---LLINTPILAQ-AHTEGVSGLTVTVVGKSRPRATSGLLHTSTA	56
Dd	1 MSRFNPSPVSLSTVLGLMFSSASAFADATKTDETM---VVTAAGYAO-----VIQNAPA	51
Qy	57 SDKIISGDITLROKAV-NLGDALGVCIHASOYGGGSAPV-IRG-----Q	100
Dd	52 SISVISREBLEGRYRDVTALKSVPGVTVT--GGDTTDISIRGMGSNYTLILVDGKRQ	109
Qy	101 TORRIKLNMHGEGTGMADFDPHPAHMVDTALSOQVEILRGPTVLTLYSSGNVAGLVADV	160
Dd	110 TSRQTRP-NSDGPGISQGWLPLQAI-----ERIEIVGPMSTLYGSDAIGVINIIT	161
Qy	161 GKIPKMPENGVSVELGLRLSSGNLEKLTSGGINIGLNKNFVLHTEGLYRKSGDYAVPRY	220
Dd	162 RKDOQQW-----SGNVQLSTVVQEN-----RASGD-----186	
Qy	221 ENLKELPDSPRFANGCHRAVLGWRRKFYRTYSDDRDQYGLPAHSHEYDDCHADIWQK	280
Dd	187 -----EOSANFFVTGPLSDALS--LQVYGQITQRDEDEI-----EHGYGDKSL-----R	228
Qy	281 SLINKRYLQLYP--HLILTSEEDVD---YDNPGLS-----CGPHDDDDAAHAH---AH	322
Dd	229 SLTSKLNQLNDHQLOLEAGVSAQDRENNNGVSKAQSOGCGTCNTDNQYRNHHVAVSH	288
Qy	323 NGKPWIDLNKRYELRAEWKQFPFGPEALRVHLNRNDY--HDEKAGDAVENFNNOQN	380
Dd	289 QG-----DWQ-----GVGQSDTYLQYEENTKSRMSIDNTVFVK	322
Qy	381 ARIELRHOPIGRLKSGWGVOYLQKSSALSATSEAVKQPMLLDNKVQHYSFFGVEQAWD	440
Dd	323 STL V---APIGEHMLSFGE--GKHESLEKTSNKISSRTHISN-TQWAGFIEDENALAE	376
Qy	441 NFTLEGGVVERVEKOKASIRYDKALIDRENYYKQPLDLAGHRQTARSFALSNWYFTQHK	500
Dd	377 QFRLTFGGLDHDK-----NY-----GSH-----FSRPVTVGMNLDPINT	411
Qy	501 LSLTASHQERLPSTQELYAHGHKVATNTTFEVGNKHLNKERSNNIELALGYE-GDRWQYNL	559
Dd	412 VKGGYSTGTFRAPQLREVTPDMQVSGGGNIYGNPDLQPETSIINKLSMTSTGSGLAASL	471
Qy	560 ALYRNRF-----GNYYIAOTLNDGRGPKSIEDDSMKLVRYNQSGADFYAGEEI	609

Query Match	5.5%;	Score	221;	DB	2;	Length	723;
Best Local Similarity	19.6%;	Pred.	No. 1.4e-07;				
Matches	157;	Conservative	130;	Mismatches	301;	Indels	212; Gaps
							41;
QY	11	LSILLINTPILAAQAHETEQQSGVLETVVCKSSPRATSGLLHTSTASDKIISGDTLROKA	70				
Db	6	LSIAITTT--LVTNALAAQSVELDSINVIAITRP---SRFAYTPKQSK---DSLLSKQA	57				
QY	71	VNLGDALDGVPGIHASQYGGG---ASAPVIRGQTGR--IKVLNHGETGMDAFSDPH--	124				
Db	58	TSVADALEIDPNVDVR--GGSRSTAQKFNIRGLSDNRVQVVDIGVRQNFDLA-----HRG	110				
QY	125	AIMVDTALSQQVEILGRPVTVLLXSSGNVAGLVADVADGKIPEKMPENGVSCELGRLSSG-	183				
Db	111	SYFPMNLQIEIENVKGFSSSLGSGALGVVAMRPNALDLLKNH---DKGVKTRQGY	167				
QY	184	NLEKLTSGGINI--GLQKNFVLHTEGLYRKSQGYAVFERYNLKRLPDSRRFANGQHRAV	241				
Db	168	OTANNLSEKDVFAANDKDPVILISGFPYNNADNLRTCKGNKLNH---TAYKQFGG---LAK	222				

```

16 INTPL-----LAQHETBOSVGLETVVVKGSRPRATSGLLHTSTASDXILS 62
      :|||:
Db    MNTPEFLRSLLSLTAAAGFAHAENNAKWLDTVTVRGDRQ-----GSKIRTNIV- 50
      :|||:
      :|||:
83 GDTLRQK---AVNLGDALDGVGIHASYQGGSAPVI-----RGOTGRRIKV----- 107
      :|||:
      :|||:
      :|||:
51 -TTQQDESTATMREHLKEEPI----DFGGNGTSEQFLTRGMGN'SVDIKVDNAYS 105
      :|||:
      :|||:
      :|||:
108 ---LNHGCTGDMADFSPDHAIWYDTLTSQQVEITLRFPTLLTYSSGNVAGLVDAQKIP 164
      :|||:
      :|||:

```

Query Match	5.5%;	Score 221;	DB 2;	Length 723;
Best Local Similarity	19.6%;	Prod. No. 1.4e-07;		
Matches 157;	Conservative 130;	Mismatches 301;	Indels 212;	Gaps 41;
Qy	11	LSILLINPTLAAAHETEQQSGVLETVVVGKSRPRATSGLLHSTASDKIISGDTLRQKA	70	
Ddb	6	LSLAITTT--LVTNALAAQSVELDSINVIAITRDP---SEFAYTPKQSK---DSLILSKQA	57	
Qy	71	VNLGDALDGVPGIHASGYGGG---ASAPVIRGQTGR--IKVLNHHGETGDMADSPDH---	124	
Ddb	58	TSVADALEDIPNDVDR--GGSRISIAQKPIRIGLSDNRRVVQVDIGVRQNFDLA-----HRG	110	
Qy	125	AIMVDTSALSQVEITLRGPVTLTYSSGNVAGLVADVADGKIPEKMPENGVSGELGLRLSSG-	183	
Ddb	111	SYFIPNSLIGIEIVIKGPSSISMGSGAIGGVAMRTPNALDLKNN-----DFGQVKIRQGY	167	
Qy	184	-NLEKLTSGGCINI--GLGKNFVLHTEGLYKRSGDYAVPYRNLKRLPDSPPRFANGQHRAV	241	
Ddb	168	OTANILSEKVDYFPAANDKFDVLISGTFYFNAADNLRTGKNKLNN---TAYKQFGG---LAK	222	

Db 106 SQUILYHQR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAI 146
 Qy 165 EKMP-----NGVSGELGRLSG--NLEKLTSGGINIGLGNFVLHTEGLY-----RKSQ 213
 Db 147 AKTVDAQDLKGLDKWGVRLNSGFASNEGVSYGASVFGKEGNF-----DGLFSYRNDEK 202
 Qy 214 DYAVPR-YRNL---KRLPDS-----RRFANGQHRVAVLGRKRFYR-----251
 Db 203 DYZAGKFRNVNGKTPVSALDKRSYLAKITGTFGDDHRIVLSHMKQHRGIRTVRE 262
 Qy 252 -TVSDRRDQYGLPAHSHVEDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNFGLSCG 310
 Db 263 FTVGDKSSRINIDROAPAYR-----TTQSTNLTAYTGKILG 299
 Qy 311 FHDDDDAHAHNGKFWIDLRNKRY---ELRAEWKQFPFGFEALRVHLNRNDYHDEKAG 367
 Db 300 FVEKLDANATV-----LEKERYGADSGTGYAGNVKGPNEHRTITTCGANFNFDRLA 351
 Qy 368 DAVENFNQOT-QNARIELRHQ---PIGRUKSGWQVYLGQKSSALSATSEAVKQPM---420
 Db 352 E-----QTLKVGGINVRHOBKIPQAFNLNSKFSIPTTEKN-----GQKVDKPEMQO 397
 Qy 421 -----LLDNKQCHYSPFGVEQA-NWDFTEGGRVVEKQKASIRYDKALIDR 466
 Db 398 MKDRADEDTVHAKLSNPTKIDTGVVVEALHDIGFTLGLRYDRFKVK-THDGKTVSS 456
 Qy 467 ENYKQPLDGAHRQARSFALSGNNYFTPOHKLISLTASHOERLPSTOELVAHGKH---523
 Db 457 SNLN-----PSFGVIWQPH-----HWSFGASHNY---ASRSPRL--YDALQTHGKRGII 501
 Qy 524 -VATNTEVGNKHLNERSNNIELALGYEGDRWQYNLALYRNFNGYIYAOTLNDGRGPK 582
 Db 502 SIADGT-----KAERANTEIGFNYNDGTFAAN-----GSYFW-QTKDALANP 544
 Qy 583 SIEDDSEMLRVYNOSGADPFGAB-GEIYFKPTPRYRIGVSGDYVRGLKALNPLSPGRED 641
 Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASVYTGGLTAKGVVS-----HSKPRFYD 593
 Qy 642 AYGNRPIAQQDQVAPRPAARLGHKLKASLTDRI-DANLDY-YRVFAQNKLYARTTRTP 699
 Db 594 THDKLLSANPEF-----GAQVGRTWASLAYRFQNFNLEIGWR-----GRYVQKAT 640
 Qy 700 GHMLNLGANYRR-----NTRYGEW-----NVVYKADNLNLSVVAHSS 738
 Db 641 G-SILAGQDKRKNLENVVRKGFVNDVFNANKPLOGKDTLNLSNVNPNFKFYYPHSQ 699
 Qy 739 FLSDT-POMGRSFTGGVNVKF 758
 Db 700 RWTNTPGVGRDRLGVNVKF 720

 RESULT 18
 F81962
 probable iron-regulated outer membrane protein NMA0453 [imported] - Neisseria meningitidis
 C:Species: Neisseria meningitidis
 C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: F81962
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jagels, K.; Leach, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: AB1775; MUID:20222556; PMID:10761919
 A:Accession: F81962
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-714 <PAR>
 A:Cross-references: GB:AL157959; NID:g7379120; PIDN:CA883750.1; PID:g737920
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: feta; NMA0453
 C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
 Query Match 5.3%; Score 213.5; DB 2; Length 714;

Best Local Similarity 22.0%; Pred. No. 4.9e-07;
 Matches 188; Conservative 103; Mismatches 308; Indels 257; Gaps 47;

 Qy 11 LSIILLNTP-LAQAHTQESVGLFTVTVCKSRPRATSGLLTSTASDKIISGDTLRQK 69
 Db 8 LSLISLTLAAGFAHAENANVALDTVTVKGDRQ-----GSKRTNIV---TLQOK 55

 Qy 70 ----AVNLGDALDGVPGIHASQYGGGASAPVI-----RGQTGRRIKV-----LNHH 111
 Db 56 DESTATDMRELLKEEPSI---DFGGNGTSGFLTRGMGQNSVDIKVDNAYSQSLVHQ 112
 Qy 112 GETGMADFPDHAIMVDTALSQQVEILRGVPTLLYSNGVAGLVVDVADGKIPKMPB--169
 Db 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDQA 153
 Qy 170 ---NGVSGELGRLSG--NLEKLTSGGINIGLGNFVLHTEGLYRKS-----GYAVPR-219
 Db 154 DLLKGLDKWGVRLNSGFASNEGVSYGASVFGKEGNF-----DGLFSYRNDEKYEAGK 209
 Qy 220 YRNL-----KRLPDS-----RRFANGQHRVAVLGRKRFYRRTYSRRDQY-GLP 263
 Db 210 FRNPNFGKTPVSALDKRSYLAKITGTFGDDHRIVL-----SHMKQHRGIR 258
 Qy 264 AHSHEYDDCHADIIWQKSLIN-KRYLQLYPHLLTEEDVDYDNFGLSCGFHDDDAHAHAH 322
 Db 259 TVREEF-----TVGGDKERISMKRQAPSRYB-TTQSTNLTAYTGKDLGFVEKLDANATV-311
 Qy 323 NGKPMWIDLRNKRY---ELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNQOT-378
 Db 312 -----LEKKRYSDSGSGVAGNVKGNPHNTQITTCGANFNFDRLAE-----QTL 356
 Qy 379 QNARIELRHQ---PIGRUKSGWQVYLGQKSSAL-----SATSEAVKQPMLLDNKVOHY 429
 Db 357 LKYGINYRHOETKQAFNLNSQFKIE--DKEKATDEEKNKRENEKIAKAYRLNPTKTD 413
 Qy 430 SPFGVEQAN-WDNFTLEGGRVVEKQKASIRYDKALIDRENYKQPLDGAHROTARSA 488
 Db 414 AGAYTEAHEIDGFTLTGGLRYDRFKVTHDGK-----TVSSSS 452
 Qy 489 LSGNW--YFTPOHKLISLTASHO--ERLPSTOE-LYAHQKH-----VATNTEVGNKHLNKE 539
 Db 453 LNPSFGVIWQPHHEHWSFASHNYSRSLRYDALQTHGKRGII SIADGT-----KAE 504
 Qy 540 RSNWIELALGYEGDRWQYNLALYRNFNGYIYAOTLNDG-RGPKSIEDDSEMK-----591
 Db 505 RARNTTEIGFNYNDGTFAAN-----GSYFW-QTKDALANPQNRHDSVAVREAVNAGY 555
 Qy 592 -----LVRYNOSGADPFGABGEIYFKPTPRYRIGVSGDYVRG---R 629
 Db 556 IKNHGYELGASVYTGGLTAKGVSHSKPRFYDTHDKLLSANPEFGAQVGRWTWASLAYR 615
 Qy 630 LKQVLSLP---GR--EDAYGNRPPIAQQDQVAPRPAARLGHKLKASLTDRIDANLDY 683
 Db 616 FXN-FNLEIGWGRYVQKAVGSILVAGQDKRSGLKLENVVRQGFVN-----660
 Qy 684 RYFAQNKLYARTTRTPGHHMLNLGANYRNRTRYGEMWYKADNLNLSVVAHSSFLSD 743
 Db 661 DYFANWKPLGKDTLN-----VNLSVN-----NVNFTFYPHSQRWNT 698
 Qy 744 -POMGRSFTGGVNVKF 758
 Db 699 LFGVGRDRLGVNVKF 714

 RESULT 19
 D81215
 TonB-dependent receptor, probable NMB0293 [imported] - Neisseria meningitidis (strain N
 C:Species: Neisseria meningitidis
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: D81215
 R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
 Hickey, E.K.; Hart, D.H.; Salzberg, S.D.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

[illegible]

```

Db      536 ----- :.:|.: |.:|.: |.:|.: |.:|.: |.:|.: |.:|.: |.:|.:
QY      625 ----- YVGR LKNLPSLPREDAYGNRFPIAQDQNAPRVPAARLGFLKASLTDRIDAN 679
Db      575 IPKKLYLRGSLGVMAQAK-----VVEDKENPDRV-----GIHL--NNTSNVTGN 615
QY      680 LDYYR-----VPAQNKLARYETRT-----PGHEMLN--LGANYRR-NTRYGE 718
Db      616 L-FFRYTPTENLYGEIGVTGKGKYGYDSRNKEVTLFGFARVDAMLGNHKNVNVTFA- 673
QY      719 WNVVVKADNLINQSIVYAHSSFLSDTPQMGRSFTGGVNVKF 758
Db      674 -----AANLFNQWRSDSM-----PGNPRGYTARVNYRF 703

RESULT 25
EB83360
conserved hypothetical protein PA2289 [imported] - Pseudomonas aeruginosa (str
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: EB8360
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportuni
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: EB8360
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-710 <STO>
A:Cross-references: GB:AB040464; GB:AE004091; NID:g9948311; PIDN:AAG05677.1; GS
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2289

Query Match          5.2%; Score 208; DB 2; Length 710;
Best Local Similarity 21.9%; Pred No. 1.2e-06;
Matches 184; Conservative 109; Mismatches 290; Indels 256; Gaps 46

QY    21 LAQHETEQSGLTEVTTVGKSRRATSGLLHT--STAASKIIISGDTLRQKAVNLGDALD 78
Db    25 IACAVQGERRSLEPMVTVGSYNESDITDLFPFSVDSIERRQIADG----QLGINLSEVL 80
QY    79 GVPG-----THASYGGGA-SAPVIRGOTGRIKVLNHGCTGMADFSPD-- 123
Db    81 RVPGLVQNRQNYAADLQISSRGYGARSFAFIRG-----LKLL-----ADGIPASTPDGQ 130
QY    124 -HALMDVTALSOQEILLRGPVTLLYSNGNAVGLDV--ADCKIKPKMPENGVSSELRL 180
Db    131 QQAATLNLDVAERLEVLRGPASTIYGS-NAGGVITQMSRDQQGAPRVGAETVGSDG--L 187
QY    181 SSGNLEKLTSGGINIGLGNFV-----LHTEGLYRKSGDYAVPR-----YRNLKRPLDSPR 231
Db    188 SRNHL--YTBGE---GPGVGFLVDASRMWDG-YR---DHSAARRDQTFFAKLNFPPADS 238
QY    232 RFANGQRHAVLGMWKRYFRITYS-----DRDDQYGLPAHSHEYD-----DCHAD 275
Db    239 RLA-----LIYSLSLEQNDIEDPLGQTWDAYKYDPYSVTANAELYDTRKS 282
QY    276 IIWQSLIN-KRY-----LQLYPHLITEEDVDYDNPLSCGFHDDDAHAHNKPMID 329
Db    283 IDHQAGMWERYFGEATLQVNAVGRSVVQYAIKQIGC-----RSNPRCQNGAVID 338
QY    330 LRNKRYELRAEWQPF---PGFEALRVHLNRNDVHHDEKAADVAFENFNNTQONARIELR 386
Db    339 FDRDFHGTYVRVLQPVSQAPGELNLTVGLD-YDQRSRDDRG--YQNFNGDQL----- 387
QY    387 HQPIGRKLSGWQYLQKSKSALSASEAVKQPMILLNKNVQHSYFFGVEQCANW--DNFTL 444
Db    388 -----GVGKILURDEVDTATS---LDPYL-----QASWAIDAWTL 419

```

```
QY 445 EGGVVEKQASIRYDKALIDRENYKQPLDPLGAHQRTARSFALSGNNYFTFPQHKLSLT 504
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 QAGVHSTMKWEV-----DDRYLSNGDASGRRYKNTSPFSVMYAFPTDLHGYS 470
QY 505 ASHQERLPSTOEL-YAHGKHVATNFEVGNKHLNKNERSNNIELALGYEGDRW--OYNLA 560
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
471 AKGGETTQEMAYA---PVANAPDVNFGLKATSSQYB--AGLKAFLMGNTRVNAA 525
QY 561 LYRNFGNYIYAQTLDNGRGPKSIEDDSEMFLVRYNQSADF-----YGAEGEI----- 609
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 IFQVTEDEIVASSLGR-----TSQNAGKTLRRGFELGLESELSSEHWN 572
QY 610 ---YFKPTPRYIGV-SGDYVRGRLKNLPSLP-----GREDAYGNRP 647
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
573 NLAYTRLSATYDSDFEAGGKTIGKGLHFGVPFESSLFGELVWKPAGISMGWEGMYRSQV 632
QY 648 FIAODQONAPRVPAAARLGHFLKASLTDRIDANLDYRVFAQNKLAARYETRTPGHMLNLG 707
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
633 YV-EDSSEKAAPS-----YAVF--NWRTRFE----- 656
QY 708 ANYRNTRYGENWY--VKADNLLNQSVYAHSSFLSDTPQM-----GRSFTGGVNVKF 758
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
657 -----QRLGAWTFHQLVRLDNLPRQ-YVGSVIVDGNRRYEEAAPGLSMVAGAVEY 708

RESULT 26
Ti2069
hemim receptor - Yersinia pestis
C:Species: Yersinia pestis
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: Ti2069
R:Horning, J.M.; Jones, H.A.; Bertolino, V.J.; Perry, R.D.
Submitted to the EMBL Data Library, October 1998
A:Description: hmuKSTUV of Yersinia pestis encodes an ABC transport system necessary for
A:Reference number: Z17401
A:Accession: Ti2069
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-676 <HOR>
A:Cross-references: EMBL:U06447; NID:g3776552; PID:g1407797
A:Experimental source: strain KIM6
C:Genetics:
A:Gene: hmuR
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
C:Keywords: membrane protein

Query Match 5.1%; Score 205.5; DB 2; Length 676;
Best Local Similarity 19.3%; Pred. No. 1.7e-06;
Matches 164; Conservative 110; Mismatches 282; Indels 293; Gaps 37;

QY 11 LSILLINT-PILAAQHETEVSQVLETVTVGKSRPRATSGLLHTSTASDK----- 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 LSLAIACLTPLATQA-----ADTTTQTSKKKSTDTMVVTATGNERSSEFAPMMVT 65
QY 60 IISGDT-LRQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTGRRIKVLNHHGDTGM- 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 VIEGNAPTSQTAATAADMLRQVPLGTVTG-----SRTNGQDVVMRGYKQGVL 115
QY 118 -----ADSPDHAIWDTALSQQVEILRGPVTLTYSSGNVAGL-----VDVADGKIP 164
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 LVDGVRQGTDTGHLNSTELDPALVKRIEIVRPAALLYGSGALGVIAETVDAAD---- 171
QY 165 EKMPEGVSGELGRLSSGNLEKLTSGGINTGLGNFVLHTE-----GLYRK 211
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 --MLQPGQNS--GYRVYS-----SAATGDHSFGLGASAFGRTRDLDGLILSGFTRDIGNIRQ 223
QY 212 SGDVAVPR-----YRNKLELPDSPRFANGQHEAVLGW 244
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 SNGFNAPNDEITSVLAKGTWQIDSIQSLSANLRYNNSALEPKNPQTSAPSSTN-VMTN 282
QY 245 RKRFRRTYSRRQYGLPAHSHEVDDCHADIIWKSILNKRYLQLYPHLLTEEDVDYDN 304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
283 RSTIQR-----DAQLRYNIKPLDQELNATAQVYSEVEINAR-----PQGSABEGREQT 333
```

```
QY 305 PGLSCGFHDDDAHAHANGKPIWIDLRNKRVELRAEWKOPFPGFALRVHL---NRNDYH 361
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 BGVK-----LENRTFLF-----IESPASHLLTYGTETYK 362
QY 362 HDEKAGDAVENP-----FNNQTONARIELRHQPIGRILKSGWGVQYLGOKSALSATSEA 415
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363 QEQTFGGATESFPQAKIRFSSGLWQDEITLRLDPLVSILAGT----- 403
QY 416 VKQPMLLDNKVQHSYFFGVGEQANWDNFTLEGGRVVEKQASIRYDKALIDRENYKQPLP 475
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
404 -----RYDNYSGSSDGY-----A 416
QY 476 DLGAHQRTARSFALSGNNYFTFPQHKLSLTASHQE--RLPSTOELYAHGKHVA----- 525
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 DVDADKWSRG-AIS---ITFTDMLFLFGSYAQAFRAFTWCEMYNDSKHFAIPIRPGLT 471
QY 526 -TNTEVGNKHLNKNERSNNIELALGYE-----GDRWQYNLALYRNFGNYIYAQTLD 577
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
472 LTN-YWVPNPNLKPETNETQYGFGLRPSDLLMAEDDLQFKVSYFDTRAKADYISTRV--- 527
QY 578 GRGPKSIEDDSEMFLVRYNQSADFYGAEGEYIFKPTPRYRIGVSGDVYRGLKNLPSLP 637
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
528 -----DMQAMTTTSSVNIQAKIWDGWDASMSYK-TALFNWDLAYNTRGKNQ----- 573
QY 638 GREDAYGNRPFTIAQDDQONAPR-----VPAARLGHFLKASLTDRIDANLDYRVFAQNK 691
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
574 --TDEM-----LDTINPDVTISVDYVPVANSFP-----SVGMGTGTA-NRS 611
QY 692 ARYETRTTP--GHHMLNLANGYRNTRYGENWYVKADNLLNQSVYAHSSFLSDTPQMGRS 749
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
612 SRVSSSTPQAGYGVNDFYVSYKQBAFKGMTTLLGNVFEKEYTPOGI-----PDGR- 666
QY 750 FTGGVNVKF 758
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
667 -----NVKRF 670

RESULT 27
AG0035
hemim receptor precursor [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AG0035
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.F.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0035
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89146.1; PID:g15978384; GSPDB:GN00175
C:Genetics:
A:Gene: hmuR
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog

Query Match 5.1%; Score 205.5; DB 2; Length 676;
Best Local Similarity 19.3%; Pred. No. 1.7e-06;
Matches 164; Conservative 110; Mismatches 282; Indels 293; Gaps 37;

QY 11 LSILLINT-PILAAQHETEVSQVLETVTVGKSRPRATSGLLHTSTASDK----- 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 LSLAIACLTPLATQA-----ADTTTQTSKKKSTDTMVVTATGNERSSEFAPMMVT 65
QY 60 IISGDT-LRQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTGRRIKVLNHHGDTGM- 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 VIEGNAPTSQTAATAADMLRQVPLGTVTG-----SRTNGQDVVMRGYKQGVL 115
QY 118 -----ADSPDHAIWDTALSQQVEILRGPVTLTYSSGNVAGL-----VDVADGKIP 164
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 LVDGVRQGTDTGHLNSTELDPALVKRIEIVRPAALLYGSGALGVIAETVDAAD---- 171
QY 165 EKMPEGVSGELGRLSSGNLEKLTSGGINTGLGNFVLHTE-----GLYRK 211
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 --MLQPGQNS--GYRVYS-----SAATGDHSFGLGASAFGRTRDLDGLILSGFTRDIGNIRQ 223
QY 212 SGDVAVPR-----YRNKLELPDSPRFANGQHEAVLGW 244
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 SNGFNAPNDEITSVLAKGTWQIDSIQSLSANLRYNNSALEPKNPQTSAPSSTN-VMTN 282
QY 245 RKRFRRTYSRRQYGLPAHSHEVDDCHADIIWKSILNKRYLQLYPHLLTEEDVDYDN 304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
283 RSTIQR-----DAQLRYNIKPLDQELNATAQVYSEVEINAR-----PQGSABEGREQT 333
```

```
Db 116 LVDGVRQGTDTGHLNTEFLDPAIVKRIEIVRGPAALLYGSGAGGVIAETVDAAD----- 171
QY 165 EKPENGVSGELGLRLSSGNLEKLTGGGINIGLGNFVLHTE-----GLYRK 211
Db 172 --MLQFGQNS--GYRVYS--SAATGDHSFGLGASAFGRDLDGLSFGTRDIGNIRQ 223
QY 212 SGDYAVPR-----YRNLKRLPDSRRFANGOHRAVLGW 244
Db 224 SNGFNAPNDETTISNLVAKGTWQIDSTQSLSANURYNNAGIEPNQTSAPSSTN-VMTN 282
QY 245 RKRFFRYTSDRRDQYGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDN 304
Db 283 RSTIQI---DAQLRYNRPDQEWLNATAQVYSEVINAR-----PGSAEEGREQIT 333
QY 305 PGLSCGCFHDDDAHAHANGKPMIDLRNKRVELRAEWKQFPFGFHALRVHL-----NRNDYH 361
Db 334 EGVK-----LENRTL-----IESPASHLLTYGTETYSK 362
QY 362 HOEKAGDAVENF-----FNNQTONARIELRHOPIGRLKSGWGVQYLGQKSSALSATSEA 415
Db 363 QEQTPGATESFPQAKIRFSSGWLQDEITLRLPVSILAGT----- 403
QY 416 VKQPMLLDNKQVHYSFSGVEQANWMDNFTLEGGVRVEKQKASIRYDKALIDRENIYKQPLP 475
Db 404 -----RYDNVSGSSDGY-----A 416
QY 476 DLGAHQRTARSFALSGNWYFTPOHKLTLTASHQE--RLPSTQELYAHGHVA----- 525
Db 417 DVDADKWSRG-AIS-----ITPTDMLFGSYAQAFRAFTMGEMVNDSKHFAIPRPGLT 471
QY 526 -TNTPEVGNKHLNKERSNNIELALGYE-----GDRWQYNLALYRNFNYIYAOTLND 577
Db 472 LTN-YWVPNPLKPEINETQYGFGLRFSDDLMAEDDLQFKVSYFTKADYISTRV----- 527
QY 578 GRGPKSIEDDSEMKLVYRQSGADFYGAGEIYFKPTPRYRIGVSGDYVGRGLKNLPPLP 637
Db 528 -----DMQAMTTTSSVNIIDQAKIWGDASMSYK-TALFNWDLAYNRTRGKNQ----- 573
QY 638 GREDAVGNRPFFAQDQONAPR-----VPAARLGFHLKASLTDRIDANLDYRVPQONKL 691
Db 574 --TDEW-----LDTINPTVTSIVDVPVANSFG-----SVGWIGTFA-NRS 611
QY 692 ARYETRTP--GHHMLNLGANYRNRTRYGBWNWYVXADNLLNSVYAHSSFLSDTPQMGRS 749
Db 612 SRVSSSTPOAGYGVNDFYYSYKQAEFGKMTTLLGNVFEKYYTPQGI-----PODGR- 666
QY 750 FTGGVNVKF 758
Db 667 -----NVKF 670

RESULT 28
A57148
outer membrane protein hmxC - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
C>Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
C:Accession: A57148
R:Cope, L.D.; Yoge, R.; Muller-Eberhard, U.; Hansen, E.J.
J. Bacteriol. 177, 2644-2653, 1995
A:Title: A gene cluster involved in the utilization of both free heme and heme:hemoexin
A:Reference number: A57148; PMID:95270579; PMID:7751272
A:Accession: A57148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <COP>
A:Cross-references: GB:U09840; NID:G595325; PIDN:AAA87059.1; PID:G595326
A:Experimental source: type b
C:Genetics:
A:Gene: hmxC
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
C:Keywords: membrane protein
F:59-189/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:408-725/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
```

Query Match 5.0%; Score 203; DB 2; Length 725;

Best Local Similarity 19.8%; Pred. No. 2.8e-06;

Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;

QY 11 LSTILLINTPLAQAHTEQSVGLTETVTVVCKSPRATSGLLHTSTASDKLIISGDTLRQKA 70

Db 6 LSLAIAT--LVTANALQSVELDSINVIAITRDP--SRPAYTEKQSK--DSLSSKQA 57

QY 71 VNLGDALDGVPGIHASQYGGG--ASAPVIRGOTGRR-IKVLNHHGTGMADFSPPH-- 124

Db 58 TSVAALALEDPNVDIR--GGSRSIAQFNIRGLSDNRVQVIVDVRQNFDA-----HRG 110

QY 125 AIMVDTALSOQVEILRGVPTLYSSNGVAGLVDAVDGKIPEKMPGVSGELGLSSG- 183

Db 111 SYPLPMSLIQEIVKIPGSSSLWGSGALGGVAMRTPNALDLDLKN--DKFGVKIRQGY 167

QY 184 -NLEKLTSGGINI-GLGKNFVLHTEGLYKRSQGYAVPRYRNLKRLPDSRRFANGOHRAV 241

Db 168 QTANNLSESDVFAANDKFDVLISGFYNNADLRTGKGNKLN--TAYKQFGG--LAK 222

QY 242 LGWRKRFYRRT-YSDRRDQYGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTEEDV 300

Db 223 FGWQINDANRVELSHRETRFKQTAPN-----NEVENELTNEQITDQIR 266

QY 301 DYDNPGLSGCFHDDDAHAHANGKPMIDLRNKRVELRAEWKQFPFGFALRYHLNRNDY 360

Db 267 EFHKP-----NNGSP-----PKAKPSQEEF-----YSGVKTRFGSVSY 299

QY 361 HHDEKAGD-AVENFF--NNQTONARIELRHOPIGRLKSGWGVQYLGQKSSALSATSEA 415

Db 300 LTDQQLPDQSTVFNYYLTDPNPLYNTHIALYNNKT-----IEKEQRKVSQVKDQTKL 351

QY 416 VKQPMLLDN--KVQHYSF-FGVEQANWMDNFTLEGGV-----RVEKQKA----- 455

Db 352 TTRGINLRSSSELSHLSFVYGVVDYMR-DKIRTERGTNNKDAQPRADPYNANSNTTGYLLI 410

QY 456 -----STRYD-----KALIDRENIYKQPLDGAHQRTARSFALSGNWYF 495

Db 411 AHIPLFGEXLLSPSVRYDHYDTSSKTVKYKDNHL-----SPATKLTIWV 455

QY 496 TPQHKLTLTASHQE--RLPSTQELYAHGHVAT-----NTFFVGNKHLNKERSNNIE 545

Db 456 T--NMLDFTAKYNEAFRAFSQMOERFVSGSHFGTSLILGRNEINKF-VANPNLRDETAKNKE 512

QY 546 L-----ALGYEGDRWQYNLALYRNFNYIYAOTLNDGRGPKSIEDDSEMKLVRYNQ 598

Db 513 ITANLHFDLSLFKQGDKFKIEATYFRNDVKDFNLKIFNDAKNTNNSASA-----GA 564

QY 599 GADFYGAGEIYFKPT-PRYR-----IGVSGDYVRGRL---KNLPSPLPGREDAYGN 645

Db 565 GA---GANPNGALLPTKSOYQNTNARLSGIELQAOYQTERLTFTFTNYGSTKCKDSDG- 620

QY 646 RPFIAQDQONAPRVPAAARLGFHLKASLTDRIDANLDYRVPQONKLARYSTRT-PGH 701

Db 621 -----EALSNIAASKIGVGNVALVKDFTVGATVTTHYA--AQRVVPKDHSHVTPSY 670

QY 702 HMLNLGANYRNRTRYGEW 719

Db 671 ILTDLRATYA--PLKGEW 686

RESULT 29

AC2075

outer membrane heme receptor alr2153 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AC2075

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguc.

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A

J. Bacteriol. 177, 2041-2049, 1995

A; Title: Cloning, sequencing, and characterization of the gene encoding FxpB, a major A; Reference number: A56268; MUID: 95238275; PMID: 7721696

A; Accession: A56268

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-713 <BEU>

A; Cross-references: GB:U13980; NID:g833694; PIDN:AAC43332.1; PID:g833695

C; Genetics:

A; Gene: ffpB

C; Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology F; 1-22/Domain: signal sequence #status predicted <Sig>

F; 62-191/Domain: tonB-dependent receptor amino-terminal homology <TNN>

Query Match 5.0%; Score 202; DB 2; Length 713;
Best Local Similarity 21.6%; Pred No. 3.2e-06;
Matches 187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;

Qy 11 LSLILNTPL-LAQAHETEQSVGLTEVTTVTKGSRPRATSGLLHTSTADKLIISGDTLRQK 69
Dl :
Db 8 LSLLSLTLAAGFAHAENANVALDVTVKGDRQ-----GSKRTNV--TLQQ 55
Qy 70 ----AVNLGDALGVPGTHASQYGGASAPVI-----RGQTGRIRKV-----LNHH 111
Dl :
Db 56 DESTAYDMRELKPEFSI---DFGGNGTSQFLTRGMQNSVDIKVNAYSQSILYHQ 112
Qy 112 GETGDMADFDPHAMIVMDLTALSQQVEILRGPVTLTYSSGNVAGLVADVAGDKIPKMPE-- 169
Dl :
Db 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDAQ 153
Qy 170 ----NGVSGELGLRLSG--NLEKLTSGGINLGKNFVLHTEGLY----RKSGDYAVPR- 219
Dl :
Db 154 DLLKGLDKRWGLNSGFAGNNGASYGVFKEGNF---DGUFYSNRNDKDYEAGKG 209
Qy 220 YRN---LKXLPDSP-----RRFANGQHRAVLGMRKRFRY--RTYSDRDQYGLP 263
Dl :
Db 210 FRNDNGKTPVSALDKRSYLAKITGTCDGDHRIVLSHMKDQHGIRTV--REEPAVS 266
Qy 264 AHSEHYDDCHADIWOKSLINRYQLPHLLITEEDVDYDNPGLSCGHDDDDDAHAHN 323
Dl :
Db 267 EKN-----SRITIQRAPSRYE-TTQSTNTLAYTGKDLGEVEKLDAVV-- 310
Qy 324 GKPWDILNKRYELRAEWKPPFGFEALRVHLNRNDYHDEKAGDAVENFNNOQNARI 383
Dl :
Db 311 -----LEKKRYS--ADKDX-----NGYAGNVKGNPHTRIATSMNFPDS 348
Qy 384 ELRHQPIGRLKSGWGVOYLQG--KSSAL-----SATSEAVKOPMLLDNKKVQHS 430
Dl :
Db 349 RLAEQTL--LK--YGINYRHQBIKPQAFNLNSEFEIKDKKEKATNEEKKNRENEKIAYR 404
Qy 431 FFCVEQANW-----DNFTLGGVVRVEKQKASIRYDKALIDRENKYKQPLDLAGH 480
Dl :
Db 405 LTNPTKTDGTGYAEIHIEDGFTLTGGUKYDRFKVKTHDGK----- 445
Qy 481 ROTARSFALSGNW--YFTPOHKLSLTASHQ--ERLPSTQE-LYAHGKH-----VATNTFEV 531
Dl :
Db 446 --TVSSSSLNPSFGVIWQPREHWSFSASHYAGSRPLRYDALQTHCKRGIISIADGT--- 500
Qy 532 GNHINKERSNNIELALGEGEDRWQYNLARYNRPCNYIYAQTLNDG-RGPSKSIDDSEM 590
Dl :
Db 501 -----KAERARTEIGFNTNDGTFAAN-----GSY-FRQTIKDALANPNQEHDSVAV 546
Qy 591 K-----LVRYNOSGADFYGAEGEITYPKTFPRYIGVSGD 624


```

Db      547 REAVNAGYIKNHGYELGASRYTGGITAKGVSHSFRFYDTHKDLKLSANPEFGAQVGT 606
Qy      625 YVRG---RLKNLPSLP-----GR--EDAYGNRPFFIAQDDQNAAPRVPAARLGHKLKASLTDR 675
Db      607 WTASLAYRFKN-PNLEIGWRGRYGVQKAVGSILAAQKDRDGLKLENVVRQGFVN-----659
Qy      676 IDANLDYRVFAQNKLAARYETTPGHHMLNLCANRYRNRTRYGEWNWYKADNLLNQSYIA 735
Db      660 -----DVFANWKLPGKDTLN-----VNLVSN-----NVFDKFTYP 689
Qy      736 HSSFSLDT-PQWGRSFTGGVNVKF 758
Db      690 HSQRWNTLPGVGRDVLGVNVKF 713

RESULT 31
AC0574
ferrienterobactin receptor precursor [imported] - Salmonella enterica subsp. enterica se
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0574
F:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
. S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0574
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-751 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05061.1; PID:gl6501837; GSPDB:GN00176
C:Genetics:
A:Gene: STY0628
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Query Match 5.0%; Score 200.5; DB 2; Length 751;
Best Local Similarity 20.2%; Pred. No. 4.4e-06;
Matches 171; Conservative 97; Mismatches 299; Indels 281; Gaps 40;
Qy      17 NTPLLAAQHAETEQQVGLTETVTV--VGKSRP-RATSGLLHTSTASDKIISGDTLQKAVNL 73
Db      38 DTIVVTRAAQNLQAPGVSTIADIRKPPARDVSEIIRT-----MPGVNL 83
Qy      74 -GDALDGVPG-----IHASQYGGGASAPVIRGQ--TGRRIKVLNHHGE---TGMADFSPD 123
Db      84 TGNSTSGQRGNRRQIDIRGMGPENTLILIDGKPVTSRNSVRLGWRGERDTRGDTAWVPE 143
Qy      124 HAINVDITALSQOVELRGPVTLTYSSGNVAGLVADVADGK-----IPEKMP 169
Db      144 -----MIRIEVLRGPAARVYNGAAGGVNIIITKKGSEWHGSMWNTYFNAPEHDE 195
Qy      170 -----NG-VSGELGLRLSSGNLEKLTSGGINIGLKNFVLHTEGLYKSGDVAVP 218
Db      196 GATKRTNFSNLGPGDPSFL-IGNLDKTDADARNINQHQ-----SERIGSYA-- 244
Qy      219 RYRNKRLPSPRRFANGQRAVIGW-----RKPFYRTYSDRDQYGLPAHSHEYDDCH 273
Db      245 -----DTLPAGREGVINKDINGVVRWDPAPIQLSLEAGYSRQGNLYAGDTQNTNTQLV 299
Qy      274 ADITWQSLINKRYLQLYP-----HLTTEEDVDYN-----PGLSCGFH----- 312
Db      300 KDNVGTET--NRLFRONYSLTWGNGWNGVNTSNWQVYEHTRNSRMPEGLAGGTGIFDP 357
Qy      313 -----DDDDAHAHNGKPIDLRNKRKYELRAEWKQ-----PFG 347
Db      358 KASQYADADLNDVTLHSEVSLPFDLLVQNQLTGTGWAQRMKQDLQNSQTFMGGINP 417
Qy      348 FEALRVHLNRNDYHDEKAGDAVENFFNNQTONARIELRHQPIGRLLKSGMGVQYLGKSS 407

```

```

Db      418 YSS-----TNRSPYSKAEIPLFAEN-----NMEL-----442
Qy      408 ALSATSEAVKQPMLLDNKNVQHYSPFGVBOANW-----DNFTLEGVVRVEKQKAS 456
Db      443 ----TDSTMLTPGI---RFDHHSIVG---DNWSPSLNLSQGLGDDFTLKWGI-----484
Qy      457 TRYDKALIDRENYKQPLPDLGAHQRTARSPALSGNWNFTTPQHKLSLTASHOERLPSTQE 516
Db      485 ----ARAYKAPSLYQTNPNYL-----YSGKQG 508
Qy      517 LYAHGKHVATNTEFVGNKHNKERSNNLELALGYEGDRQYNLYALNRPFGNYIYA---572
Db      509 CYATGAGTGICCYMMGNDLKAETSKINKEIGLEFKRDGMLAGVTFWRDNRKNIKAGTVP 568
Qy      573 -QTLNDGRG-----PKSIEDDSEMKLVRYNQSGADFYGAEGIYF-----KPT---614
Db      569 LQRIINNGKTDVYQWENVPKAVVEGLEGL---NVFVSQTVNNTVNNVYMLQSKNKEGTGR 625
Qy      615 ----PRYIGVSGDYVRGRKLNLPRL-----PGRSDAYGNRPFFIAQDDQNAAPRVPA 662
Db      626 LSIIPOYTINSTLSWQVRQDVSLQSTFTWYKQEPKYDYQGN-FVTGTDKQAVS--PVS 682
Qy      663 RLGFHL-----KASLTDRIDANLDYRVFAQNKLAARYETTPGHHMLNLCG-NYRNTR 715
Db      683 IVGLSATWDVTNNVSLTGGVD-NLFDKRLWRGNAQTVRDQTGTGAYMAGAGAYTYNEPGR 741
Qy      716 YGEWNWYV 723
Db      742 ---TWYM 745

RESULT 32
H81030
TonB-dependent receptor NMB1882 [imported] - Neisseria meningitidis (strain MC58 serogr
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81030
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: H81030
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <TET>
A:Cross-references: GB:AB002538; GB:AB002098; NID:g7227136; PIDN:AAF42216.1; PID:g72271
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1882
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom

Query Match 5.0%; Score 200; DB 2; Length 725;
Best Local Similarity 21.7%; Pred. No. 4.6e-06;
Matches 178; Conservative 105; Mismatches 294; Indels 244; Gaps 43;
Qy      19 PLLAQAH-----ETEQSVGLTETTVVVGKSRPRATSGLLHTST-----AS 57
Db      15 PVYQAADVSDDDPKPQESTELPTITVADRASSNDGYTVSGTHTPLGLPMTLREIPQS 74
Qy      58 DKIIISGTLR-QKAVNLGDALDGVFGIHASQYG---GGASAPVIRGQTGRRIKVLNHHGE 113
Db      75 VSVITSQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNVLFARGS-----RIANYQIN 129
Qy      114 TGMADSPDPAHIMVDTALSQOVELRGPVTLTYSSGNVAGLVADVADGKIPEKMPENGVS 173
Db      130 GIFVADALADTG-NANTAAYERVEVVRGVAGLLDGTGEPSTATVNLVRKLRTRK-PLFEVR 187
Qy      174 GELGLRLSSGNLEKLTSGGINIGLKNFVLHTEGLYKSGDYAVPYRNKLRLPDPSRRF 233
Db      188 AEGNRKHFQ-LDADVSGSLN-----TEGTLR-----GRIVST---F 220

```

```

QY 234 ANQHRAVLGWRKRYRTYSDRDYGLPAHSHEYD-----DCHADIWQKSLINKRYL 288
DQ 221 GRGD-----SWR-----RRSRDAELYGI-----LEYDIAPOTRVHAGMDYQQA----- 260
QY 289 QLYPHLLTREDVD-----YDNPGLSGCGFHDDDDAHAAHNGK-----PWIDLR-NK 333
DQ 261 -----KETADAPLSYAVDSQGYATAPGPKDNPATWNSRRHALLNLFAGIEHRFNO 312
QY 334 RYELRAEW-----KQPPFGFALRVHLNRNDYHHDKAGDAVENFF--NNQTONARIE 384
DQ 313 DWKLKAEYDYTRGRFQYGVAGVLSI-----DHTAATDLIPGYHADPRTHSASVS 365
QY 385 L--RHQPIGR-----LKGSMGVQVILGQKSSALSATSEAV-----KQPMLLDNKY 426
DQ 366 LIGKYLRFREHDLIAGNGYKVASNKYGBRSIIPNAIPNAYFSTGYPQASPAQTI 425
QY 427 QHY-----SFFGVEQANNDFTLEGGRVVEKQASIRYDKALIDRENYIKQPLDPLGA 479
DQ 426 PQYTRRQIGGYLATRPAAADNLSLILGGYVTRVGTG-SYDSR-----TQGMTYVSA 476
QY 480 HRQTARS-----PALSQNWYFPPQHKLSLTASHQERLSTQELYAHGKHVATNTEFVGNKHL 536
DQ 477 NRFTPYTGIVFDLGN-----LSLYGSIYSLFVPOKORDEHGSYLKPVV----- 520
QY 537 NKERSNNIELALGYEGD-----RWQYNIALYRNFRGNYIYAQTLLNDGRGPKSIEDDSEMKL 592
DQ 521 -----GNNLE--AGIKGEWLEGRIMASAAVYRARKNNLATAA-----GRDP----- 559
QY 593 VRYNQSGADFY-----GAEGEYFKPTPRYRI--GVSGDYVR--CRLKNLPSLPG 638
DQ 560 -----SGNTYRAANQAKTHGEWEVGGRIITPEQIOAGYSQSKTRDQDSRLNPDSVPE 614
QY 639 RE-----DAYGNRPETAQDDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLARY 694
DQ 615 RSFKLFYAY-----HPAPEAPS--GWTIGAGVRWQSEHTDTPALIRIPNPAKA 661
QY 695 -----ETTPGHHMLNLANYRNTRYGENWYVYKADNLLNQ 731
DQ 662 RAADNSRQKAVAVADIARYFRNPRA-----ELSLNVDNLFNK 699

RESULT 33
E83485
probable tonB-dependent receptor PA1271 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83485
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: E83485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-616 <STO>
A:Cross-references: GB:AB004557; GB:AB004091; NID:99947204; PIDN:AAQ04660.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1271
C:Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to
Query Match 4.9%; Score 199; DB 2; Length 616;
Best Local Similarity 22.4%; Pred. No. 4.2e-06;
Matches 155; Conservative 88; Mismatches 249; Indels 199; Gaps 39;

QY 38 VYGRPRATSGLLTSTASDKIISGDTL-RQKAVNLGDAIDGVFGIHASQYGG-GASAP 95
DQ 31 VVTRTRTACTAS---QSLAAVSVIDREDIERSQARSVPPELLRQVPGVSLANNNGGFGKNTT 87
QY 96 V-IRQGTGRIRKVLNHHGETGD-MADFPDPAHIMVDRLASQOVELRGPVTLIYSSGNA 153

```

```

DQ 88 LFLRGTSRDLVLIDGKIVGSASAGITARQDLPEV-----LIERIEVVRGPRSSLYGSAIG 145
QY 154 GLVDV-----ADGKIPEKMPENG-----VSGELGLRSLSSNLEKLTSGGINIGLGNFVL 203
DQ 146 GVIQIFTRRGDQGAQKFFSAGYGTHTQTLGSGAGVSGAGN-----GWYSLGYSS--F 196
QY 204 HTEGLYRK--SGDYAVPR--YRNKLKLPDPSRPFANGQRAVLGWRKRFYRRYTSYDRRD 258
DQ 197 DTAGINTKRACTAGTGYEDRDGYRNLSGNLRCGYRFDNGL-----ELD 238
QY 259 QYGLPAHSHEYDDCHADIWQKSLINKRYLYLYPHLLTEEDVDYDNPGLSCGPHDDDDAH 318
DQ 239 GTLLRAKSHN-----DYDQVFGNSGFANADGE 266
QY 319 AHAHNGK-----PW-IDLENKRYELRAEWKQP---PPGFEALRVHLN-RNDYHDEKA 366
DQ 267 QNLVGGRAARFTFPDPFWLVTQAGRSEDKADAYQDGRFYSRFTRDRSDLSWQNDLTLEA-- 324
QY 367 GDAVENFFNNQTONARIELRHQPIGRKLGSMGVQVILGQKSSALSATSEAVKQPMLLDNKY 426
DQ 325 GHVLTGLYDQKDEISSEAFSVDSRLNKGWFAQYLGQYG-----RQDWQLSLRR 374
QY 427 QHYSFFGYE---QAWNDNFTLEGGRVVEKQASIRYD---KALIDRENYIKQ-PLDPLGA 479
DQ 375 DNNQOFGVHDTGSAAW-GYALSDALRF-----TVSYGTAFKAPTFNELYYPDYGNDPLDA 428
QY 480 HRQTARSFALSNNWYFPPQHKLSLTASHQERLSTQELYAHG-KHVATNTEF-----VG 532
DQ 429 ETSRSLEVLGSG-----THGWGHWA VNAFRTNVDDLLG 461
QY 533 N-----KHLNKRNNIELALGYEGDRQWYNL-ALY---NRRGNVIYAQTL 575
DQ 462 NDRPAPRPMQPNNDIEARINGVELGSGWLGMDNANATFLDPQNRSG-----V 515
QY 576 NDG-----RGPKSIEDDSEMKLVRYNQSGADFYGAEGRIYFKPTPRYRIG--VSGD 624
DQ 516 NDGNELPRARWENLEIDRRFERLS-LGASVH-AEGRRYDDPANKVGLGYATLIDLRSE 573
QY 625 Y-----VRGRKLNPLSPGREDAYG-NRP 647
DQ 574 YRLNDEWLQGRIANLFG-ADHETAYGYNQ 603

RESULT 34
D81976
probable ferric siderophore receptor protein NMA0575 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81976
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandrea
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: D81976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <PAR>
A:Cross-references: GB:ALJ62753; GB:ALJ57959; NID:g7379120; PIDN:CAB83866.1; PID:g7379:
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0575
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom
Query Match 4.8%; Score 192; DB 2; Length 725;
Best Local Similarity 21.2%; Pred. No. 1.7e-05;
Matches 174; Conservative 108; Mismatches 295; Indels 244; Gaps 42;

QY 19 PLLAAQH-----BTEQSVGLTETVVGKSRPRATSGLLHTST-----AS 57
DQ 15 PVYQAQDVSVDDPKPQRSTELPTITVADRATSSNDGYTVSGTHTPLGLPMTLREIPQS 74

```

Query Match	4.7%;	Score 191;	DB 2;	Length 625;
Best local Similarity	21.4%;	Pred. No. 1.6e-05;		
Matches 128;	Conservative	81;	Mismatches 248;	Indels 142; Gaps 28;
QY	5	TLKPIVLSILLINPLLAQAHTESVGLETVTVVVKSRPRATSGLLHSTASKIISGD	64	
Ddb	2	TIKKYTLTALSVPFSGWAQQNNITDNDENVVTANRFPPQKSVL---	APVDVWTRAD	58
QY	65	TLRQKAVNLGADLGVPGGIHASOYG--GGASAPVIRGTGRRIKV-----LNHHGTGD	116	
Ddb	59	IDRWQSTNINDVRULPFGVDIADQMGQRSLFIRCTNSHWLVLDIVPLNQAGITG-	117	
QY	117	MADFSPDHAIIMDTALSSQVEILRGFTVLLYSNGNVAGLVDAVGKIPEKMPENGVS	GEL	176
Ddb	118	ASDLSS-----QIPISLVORIEYRPSRAVYGSDAIGVINILTGR---DRPGTTLSA	GL	169
QY	177	GLRLSSGNLEKLTSGGINTIGLGNFVLHTEGLYRKSGDYAVPYENLAKLPDS--	PRFFA	234
Ddb	170	G-----SNGVOTYDGSTQOKLGEDTTVTLAGNYTYSKGVDV-----VACMPCAGCP	QPD	219
QY	235	NGQHRAVLGRKRFRFYRTYSDDRDQYGLPAHSHEYDDCHADIWOKSLINKRYLIQLYP	LPHL	294
Ddb	220	RDFGMFKMLWA-----GL---EHQFNEQF-----NGFARVY----	247	
QY	295	LTEEDVDYNPLSCGFHDHDDAHAAHNGKPWIDLR---NKRYELAEAWKQPFPGFEAL	351	
Ddb	248	-----GFDN-----RSYDGYTNYSPLALIDTRKLSRSTYTGLRYKNXGIVASOFI	294	
QY	352	RVLNRNDYH-----HDEKAG-DAVENFFNNQTONARIELRHOPICRLKSGMWQVYL	G	403
Ddb	295	ASYNRTXDYNSPLFGQHDITASLDEAQ-YNLQNCNT-FOLTNGWIS-AGADMQEORTE	351	
QY	404	KQSALSATSSEAVKQPMLLDNKKVOHYIS-PFGVEQANWNFTLEGGVURVEKQKASIRYDKA	462	
Ddb	352	RKSSNQNTTADF-----TQHNHTGYLTGQQQISDVLTLEGAVSRSDNS-----	393	
QY	463	LIDRENYTKPLPDLGAHRQTARSFALSGNWYTFPOHKLSLTASHOEPLSPSTOEYL-AHG	521	
Ddb	394	-----QFGWH-----STWQTSAGWEFIDGYRLIGSYGTAYKAPNLMQLIAYS	G	436
QY	522	KHVATNTFEVGNKHINKERSNNTELALGYEG--DRWQYNLALYRRNFNGNYIIYAOTLND	G	578
Ddb	437	-----GNANLKPEEKSKOWE---GGVEGHTGPULTWRLSAYRNIDIOLIDYSNL	TNG	483

K; Hayashida, T.; Makino, K.; Yasunaga, N.; Takayama, H.
gawara, N.; Yasunaga, N.; Ohtsuji, M.; Kurakawa, A.; Isii, K.; Tokoyama, Y.
DNA Res. 8, 11-22, 2001

A; Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7
A; Reference number: A96629; UID:21156231; PMID:11258796

A; Accession: D91176

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-660 <HAY>

A; Cross-references: GB:BAE37803.1; PID:g13363854; GSPDB:GN00154

A; Experimental source: Strain O157:H7, substrain RMD 050952

C; Genets:

A; Gene: ECs4380

C; Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal

Query Match 4.7%; Score 188; DB 2; Length 660;
Best Local Similarity 19.8%; Pred. No. 2.8e+05;
Matches 157; Conservative 96; Mismatches 281; Indels 260; Gaps 34

Qy 4 TWLKPVLVSLILNT-PLLQAHTETGVSGLFTVVVGKSR-----PRATSGLLHTSTAS 57
|::||::| |::||::| |::||::| |::||::| |::||::| |::||::| |::||::|
|::||::| |::||::| |::||::| |::||::| |::||::| |::||::| |::||::|

```

7  TSLRLSLLALAVSATLPTFAFATET-----MTVTATGNARSSFEAPMWVS-VIDTSAPE 59
58  DKIIISGPTLRKAVNLGDALDGVFCIHASQYGGGASAPV-IRGQTGRRIKVL-----NHHG 112
60  N-----QTATSATDLLEHVFGITLDGTGRINGQDVNMVRGYDHRGVLVDVGRQGT 110
113  ETGMDAFSPDHAIWDTALSQQVEILRGVPVTLTYSSGNVAGLV--DVAQGIPEKMPEN 170
111  DTGHL-----NGTFLDPALIKRVEIVRGSPSALLYSGALGGVISTYDVAK---DLLQE 161
171  GVSGBELGURL--SSGNLEKLTSGGINIGLKNFVLHTRGL-----YRKSGDYA 216
162  QOSS--GFRVFTGG-----TGDHSLGLGASAFORTENLDGIVAMSSRDGRDLURNGSET 214
217  VPR-----YRNLEKRLFDSPRRFANQHRHRAVLGWRKRFY 249
215  APDESIINMLAKGTWQIDSAQSLSGLVRYYNNDAREPKNPQTVGASESSNPWVDRSTIQ 274
250  RRTYSRRDQYGLPAHSHYDDCHADIWQXSLINKLYLQLYPHLLTEEDVDVNDPGLSC 309
275  R-----DAQLSYKLAPOQNDMLNADAKIYWSEVRINAQ-----NTGSSG 313
310  GFHDDDDAHAAHAKGKWIIDLNRKRYELRABWKQFPFGFEALRVHLNRNDYHHDKAGDA 369
314  EYRQIYTKGAELEN-----RSTLFAD-----SFASHLLTYGGEYRQEHQHPGA 357
370  VENF-----FNNQTONARIELRHQPIGRUKGSWGOYLQOKSALSATSSEAVKQPMLLD 423
358  TTGFPOAKIDPSSGWLQDEITLRLPL-----384
424  NKVQHYEFFGVEQANWDNFTLEGVVRVEKQKASIRYDKALIDRENNYKQPLDPLGAHRQT 483
385  -----TLGGTRYDSYRGS-----SDGYK-----DVDADKWS 411
484  ARSFALSGNMYFTPOHKLSLTASHOE--RLPSTOELYAHGKHVATNTFE-----VGNKHLN 537
412  SRA-----GMTINTNMLWLFSGYAQAFRAPTWCMEYNDKSKHPSIGRFYTYWVPNPLR 466
538  KERNNNIELALG-----YEGDRWQYNLALYRNRFGNYTYAQTLDNDRGFKFIEDDSEM 590
467  PETNETQYGFGLRFDLMLNSDALEFKASFYDTKANDYI-----SITVDFAAA 515
591  KLVRYNOSGADFYCAEGEIVPK-PTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFI 649
516  TTMSYNVPNAKIWG--WDVMTKYTTDLFSLDVAYNTRGRK-----DSTDGEYTSINPDT 568
650  AQDDONAPRYAARLGHPLKASLTDRIDANLDYVYRVAQNKLARYETRTPGHEMLNLGAN 709
569  VTSTLN--IPIAHSGF-----SVGWGCTFADRS-----THISS 600
710  YRRNTRYGWNWYV 723
601  YSKOPGYGVNDFYV 614

```

RESULT 37
E86022
outer membrane heme/hemoglobin receptor [imported] - Escherichia coli (strain O157:H7, s
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E86022
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotberg, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference numbers: A85480; MUID:21074935; PMID:11206551
A:Accession: E86022
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-660 <STO>
A:Cross-references: GB:AE005174; NID:g12518206; PIDN:AAGS8641.1; GSPDB:GN00145; UWGP:Z49
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:

A;Gene: chuA
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology

Query Match 4.7%; Score 188; DB 2; Length 660;
Best Local Similarity 19.8%; Pred. No. 2.8e-05;
Matches 157; Conservative 96; Mismatches 281; Indels 260; Gaps 34;

QY 4 TLLKPIVLISILLINT-PLLAQAHETEOSVGLTIVTVGKSR-----PRATSGLLHTSTAS 57
DB :
7 TSLRLSLLALAVSATLPTFAFATET-----MTVTATGNARSSFEAPMWVS-VIDTSAPE 59
QY 58 DKIIISGDTLRQXAVNLGDALDGVCIHASQVGGSASAPV-IRGTGTRRIKVL-----NHHG 112
DB :
60 N-----QTATSATDLLRHVFEGITLDGTGRTNQGDDVMRMGYDHRRGLVLDVQRQGT 110
QY 113 ETGDMAFDSPDHATWTDLTALSQQVEILRGTPVTLLYSNGNVAGLV--DVADGKIPEKPEN 170
DB :
111 DTGHL-----NGTFLOPALIKRVEIYVRGPSALLYGSGALGGVISYDVAK---DLIQE 161
QY 171 GVSSELGLRL-SSGNLEKLTSGGINIGLNKNFVLHTEGL-----YRKSGDYA 216
DB :
162 QOSS--GFRVFGTGG---TGDSLSGLGASAFGRTENLDGIWAWSRDRGDLRQSNGET 214
QY 217 VPR-----YRNKKLPDSPRRFPANQHRAVLGWRKRFF 249
DB :
215 APNDESINNMLAKGTWQIDSQAQSLGSLRVYYNNDAAREPKNPQTVGASESNPMVDRTSIQ 274
QY 250 RRTYSRRDOVGLPAHSHEYDDCHADIWKQSLINKRYLQLYPHLLTEEDVDYDNPGISC 309
DB :
275 R---DAQLSYKLAPOQNDWLNAKAIYWESEVINAQ-----NTGSSG 313
QY 310 GFHDHDDAHAAHNKGFWIDLNRNKRYELRAEWKQPFPGFEALRVHLNRNDYHHDEKAGDA 369
DB :
314 EYREQITKGARLEN-----RSTLFAD-----SFASHLLTYGGEYVRQEHFGGA 357
QY 370 VENP-----FNNOQNARIELRHOPIGRLKSGWGVOYLQOKSALSATSSEAVKQPMLLD 423
DB :
358 TTGFPOAKIDFSSGWLOWDEITLRDLPI----- 384
QY 424 NKVQHYFFGVEQANWDNFLEGGVRVEKOKASIRYDKALIDRENYYKQLPDLGARHOT 483
DB :
385 -----TLGGTRVDYVRGS-----SDGYK----DVDADKWS 411
C 484 ARSPALSGNWYFTPPQHKLSLTASHQE--RLPSQTELVAHGKHVATNTFE----VGNKHLN 537
DB :
412 SRA-----GMTIPTNWLMFLGSAQAFAAPTGMCEMYNDSKFPSIGRFYNYVPNPCLR 466
QY 538 KERSNNIELALG-----YEGDRWQYNLYARNRFGNIYYAOTLNDGRGPKSIEDDSEM 590
DB :
467 PETNETOEYGFGLRFDDLMLSDALEFKASYFDTKAKDYI-----STTVDFAAA 515
QY 591 KLVRYNQSGADFYGAGEIEYFK-PTPRYRIGVSGDYVRGLKPLPSLPGREDAYGNRPFI 649
DB :
516 TTWSYNVNPAKIWG--WDVMKTITDLFSLDVAINRTRGK-----DPTDGTEYISSINPDT 568
QY 650 AQDDQNAAPRPAARLGFHLKASLTDIRIDANLDYRVFAQNKLARJETRTTGHMHLNLGAN 709
DB :
569 VTSTLN--IPIAHSGF-----SVGWVGIFADRS-----THISSS 600
QY 710 YRNTRYGEWNWTV 723
DB :
601 YSKQPGYGVNDFYV 614

RESULT 38
AC3334
metal chelate outer membrane receptor [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 23-Dec-2002
C;Accession: AC3334
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
. ; Mazur, M.; Gottesman, B.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of *Escherichia coli* K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64984
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-663 <BLAT>
A;Cross-references: GB:AE000304; GB:U00096; NID:gl788470; PIDN:AACT5216.1; PID:gl788478
A;Experimental source: strain K-12, substrain MG1655
R;Nau, C.D.; Konisky, J.
J. Bacteriol. 171, 1041-1047, 1989
A>Title: Evolutionary relationship between the TonB-dependent outer membrane transport
A;Reference number: A32056; MUID:89123100; PMID:2644220
A;Accession: A32056
A;Molecule type: DNA
A;Residues: 1-146, 'RCARCSEVHQ', 159, 'NRSEMV', 166, 'YRPRYHSGTSRSR', 182-527, 'N', 529-613
A;Note: the authors translated the codon AAT for residue 528 as Ile
R;Nau, C.D.; Konisky, J.
J. Bacteriol. 171, 4530, 1989
A;Reference number: A33868
A;Contents: corrections
A;Accession: A33868
A;Molecule type: DNA
A;Residues: 144-184;610-663 <NA2>
R;Griggs, D.W.; Tharp, B.B.; Konisky, J.
J. Bacteriol. 169, 5343-5352, 1987
A>Title: Cloning and promoter identification of the iron-regulated *cir* gene of *Escherichia coli* K-12.
A;Reference number: A28377; MUID:88058737; PMID:3316180
A;Accession: A28377
A;Molecule type: DNA
A;Residues: 1-59 <GR1>
A;Cross-references: GB:M19295; NID:g145545; PIDN:AAA23581.1; PID:g551794
A;Note: residues 26-45 were confirmed by protein sequencing
R;Steffes, C.; Ellis, J.; Wu, J.; Rosen, B.P.
J. Bacteriol. 174, 3242-3249, 1992
A>Title: The *lypS* gene encodes the lysine-specific permease.
A;Reference number: A41871; MUID:92250419; PMID:1315732
A;Accession: C41871
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-96, 'D', 98-125 <STE>
A;Cross-references: GB:M89774; NID:g466776; PIDN:AAA17054.1; PID:g466779; EMBL:X65029
R;Griggs, D.W.; Kafka, K.; Nau, C.D.; Konisky, J.
J. Bacteriol. 172, 3529-3533, 1990
A>Title: Activation of expression of the *Escherichia coli* *cir* gene by an iron-independent promoter.
A;Reference number: A35408; MUID:90284362; PMID:2160948
A;Accession: A35408
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-21 <GR2>
C;Comment: This outer membrane receptor for colicins Ia and Ib is regulated by both *cir* and *tonB* transport proteins.
C;Genetics:
A;Gene: *cir*; *cirA*; *feuA*
A;Map position: 43 min
C;Superfamily: ferriterochelin receptor; tonB-dependent receptor amino-terminal homol
C;Keywords: iron transport; membrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-624/Product: colicin I receptor #status experimental <MAT>
F;66-211/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F;365-663/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 4.6%; Score 187.5; DB 1; Length 663;
Best Local Similarity 20.0%; Pred. No. 3.1e-05;
Matches 172; Conservative 110; Mismatches 271; Indels 307; Gaps 42;

QY 6 LKPIV-----LSILLINTPLAQAHETEQSVGLFTVTVGKSRPRATSGILHTSTASDKI 60
DB 4 LNPFFVRVGLCLSAICAWPVLAVDDG-----ETMVVTASS-----VEQNLKDAPASISV 53
QY 61 IGSQDTLRQKAV-NLGDALDGVGPIGTHASQYGGGASAPVIRG-----QTGRRIK----- 106

Db 54 ITQEDLQKRPVQNLKDVLEKVEPGVQLTNEGNRKGVSIRGLDSSYTLILVDGKRVNSRNA 113

Qy 107 VLNHGETGDMADSPDHAIMVDTSALSOQVEILRGPTVLLYSSGNVAGLVADQKIKEP 166

Db 114 VFRN-----DFDLW-IPVDSI--BRIEVRGPMSSLVGSDALGVVMIITKIQCK 163

Qy 167 MPENGVSSELGL-----RLSSGNLEKLTSGGINTGLGNFVILHTTEGLYRKSGDYAV 217

Db 164 W-----SGTVVDTTIOEHRDRGDTYNGQFFTSGLIDGV-----LGM 201

Qy 218 PRYRNL-KELPDSPRFA--NGQRAVLGWRKRYRYTSDRRQYQ-LPAHSHEY--- 269

Db 202 KAYGLAKREKDDPONSTTTTGTGETPRIEGFSR-----DGNVEFAWTEPNQNHDFTAG 254

Qy 270 -----DCHADLIWOKSLINKRY-----LQLYPHLLTEBDVDVDPGLSCG 310

Db 255 YGFDEQDRDSLSLKNRLERQNYSYSHNGRWDTGTSLEKYY-----GKVENKPNPNSP 309

Qy 311 FHDDDDAHAHNGKFWIDLK--NKRYELRAEWKOPFPFGFEALRVHLNRNDYHHDEKAGD 368

Db 310 I-----TSESNTVDGKYTLPLTAINOFLTVGGEWR-----HD-KLSD 345

Qy 369 AVENFFNNQTNARIELRHQPIRLGLKSGWGYVLGOKSSALSATSEAVKQPMLLDNKYOH 428

Db 346 AV-----NLT-----GTSKTSAS-----QY 362

Qy 429 YSFFGVEQANMDNFTLEGGVRVEKOKASIRYDKALIDRENYKQPLDGAHRQTARSA 488

Db 363 ALFVEDEWRIFEPLATTGVMD-----DHETGEHWSPRAYLVYNAITDVT 409

Qy 489 LSGNWFYFPQHKLSUTASHQERLPSTQELYAHGKHVANTPE-----VGNKHLNKERN 543

Db 410 VKGW-----ATAF---KAPSLQL---SPDWTNSCRGACKIVGSPDLKPETSES 454

Qy 544 IELALYGGDR-W-----QYNLALYRN-----RFGNVIYAQTLNDGRG 580

Db 455 WELGLIYNGEGLWLGVESSVTVFRNDVKDRISISRTSDVNAAPGYQNFVGFETGANGRR 514

Qy 581 PKSIDDEMKLVRYNQSGADPYGAEIYFKPTPRYRIGVSGDYVRGLKNLPLSGRE 640

Db 515 IPVFS-----YNNVKARIQGVETELKIPFNDEWKLSINYTYNDGR----- 555

Qy 641 DAYGNRPFIADQDQNAVRPAARLGFHLKASITDRIDANLDYRVF-----AQNKLARY 694

Db 556 -----DVSNGENKPLSDLPFHTANGTLDWKPLALEDWSVSGHYTGQKRADS 604

Qy 695 ETRTPGHMLNLGANYRNRTRYGEWN-----WYVKAD-----NL-----LNOSVYAHSS 738

Db 605 TAKTPGGTYI-----WNTGNAWQVTKDKVLRAGVLNLGDKLSDRDDYSYN- 649

Qy 739 FLSDTFQMGSRFTGGVNVKPF 758

Db 650 -----EDGRRYFMAVDYRF 663

RESULT 40

AB0124

probable TonB-dependent outer membrane receptor YPO1011 [imported] - Yersinia pestis (st

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AB0124

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; PMID:21470413; PMID:11586360

A:Accession: AB0124

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-690 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC89853.1; PID:g15979079; GSPDB:GN00175

C:Genetics:

A:Gene: YPO1011

Query Match

Best Local Similarity 4.6%; Score 187.5; DB 2; Length 690;

Matches 169; Conservative 127; Mismatches 299; Indels 241; Gaps 43;

Qy 14 LLINTPLLA--QAHETEQSVGLTVTVVVGKSRPRATSGLLHTSTASDKIISG-DTLRQKA 70

Db 13 VLISLPCLAWSQSHNSKD-ELDITITVAQK-----INQQQKTPISISVLITGFDLERENI 67

Qy 71 VNLGDALDVGPIIHASQYGGASAP--VIRGOTGRRIKVLNHHGETG-DMADFSFD--HA 125

Db 68 ENIVESIMRIENVMVMKAGNPSDAGFTTMRGTT-----PGMEGIQSVGFFIDGVYA 118

Qy 126 IMVDTAL--SQQVELRGPTVLLYSSGNVAGLVADQKIPEKMPENGVSSELGLRLSSG 183

Db 119 NTFDELLDVRIEVLIRGQATLYGRNTEGSGVINVT-KDPEFSPEY---KIGUSYGY 173

Qy 184 NLEKITS---GGINIGLGKNF-----VLHTEGLYKSGDYAVPRYRNLKRLPDSPRRFAN 235

Db 174 NRTQVTVTLGGSINDSEQFSYRAALKYLYGNGYFKRDYD---GKNVNDNLNDFSGRPKL 229

Qy 236 GQRAVLGWRKRFRTYTSRRDQYGLPAHS-----HEYDDCH--ADIHWKSLI 283

Db 230 RWQPMDDGMD--VMTTFDIQNRNGNTSFTALDKIKSGQKYVDSNYIGKSDVDAYKGQV 286

Qy 284 NKRYLQLYPHLLTEBDVDVDPGLSCGFHDDDDAHAAHANGKPWD----- 329

Db 287 NAVY-----TFDDIDFTS-----VSAYVDERKVDNQDLDTRLSISEL 324

Qy 330 LRNKR-----YELRAEWKOPFPFGFEALRVHLNRNDYHHDEKAGDAVENFFNNQTNAR 382

Db 325 LMNRKTKQFSQEFRLNSKYSGPFNWL-----IGSYFYQDDENE 363

Qy 383 IELRHQP--IGRLKSGWGYVLGOKSSALSATSEAVKQPMLLDNKYHSPFG-VEQANW 439

Db 364 IDFRILPYNLQALRKS-----DIKNNYAVFGNYYNYLL 397

Qy 440 DNFTLEGGVRVEKOKASIRYDKALIDRENYKQPLDGAHRQTARSA-----LSGNWY 494

Db 398 NDVELVAGARYDYEKKLNF---LMDNGFNYPQY-----SHDNNNSGAFPLPKVGLNYY 450

Qy 495 FTPQHLK--SLTASHQERLPST--QELYAHGKHVATNTFEVGNKHLNKERSNNIELALG 549

Db 451 ITGDAMLYTSIARGYKSGCFNTLGPQSSRAYNAEYMT-TYEAGVKTWFDRIT----- 501

Qy 550 YEGDRWQYNLALYRNFQNYIYAQTLNDGRGPKSIEDDSEMKL/RYNOSGADFYGAEHI 609

Db 502 ---VRWNTSL-----FWNDMKDQOQVEVAYYPISYSVNSGKSLSR-----GLESEL 543

Qy 610 YFKPTPRYRIGVSGDYVRGLKNLPLSGREDAY-----GNRPFIADQDQNAVRPAAR 663

Db 544 AWRIITRGLTVSANVGYTDAFYKNPTEIKVDNNYIPVNYKGNRP-----ANSP----- 591

Qy 664 LGFHLKASLTDRIIDANLDYRVFAQNKLARVET-----RTPGHMLNLGANYR 711

Db 592 -----GYTYSIGADYNFLNGYFVN--ATYNVKGSTYLDNANSKKQPAYGLLDLTAGY- 641

Qy 712 RNTRYGEMWVVKADNLLNQSVYAHSSFLSTPQMGH-----SFTGGVNVKPF 758

Db 642 ENKDYGV-NWMIK--NILDETYVTRAFKMDDDGIWYGRAGEPINFQGVNFNVKPF 690

Search completed: November 14, 2003, 11:00:39

Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:01:18 ; Search time 17 Seconds
(without alignments)
2096.838 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTTLKPLVLSILLINTEPL.....FLSDTPQMGRSFTGGVNVKF 758

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	330	1 PE42 ARATH	Q95b81 arabidopsis
2	9	1.2	1980	1 MY98 RAT	Q63358 rattus norv
3	9	1.2	2114	1 MY98 MOUSE	Q94y06 mus musculu
4	9	1.2	2158	1 MY98 HUMAN	Q13459 homo sapien
5	8	1.1	196	1 YO96 AGRT5	P58792 agrobacteri
6	8	1.1	316	1 BLA3 BACCE	P06548 bacillus ce
7	8	1.1	400	1 FDH PSESR	P33160 pseudomonas
8	8	1.1	933	1 RGA4 SCHPO	O74360 schizosacch
9	8	1.1	1009	1 RGA2 YEAST	Q06407 saccharomyc
10	7	0.9	117	1 NUSM DROME	P18930 drosophila
11	7	0.9	117	1 NUSM DROSA	F51940 drosophila
12	7	0.9	117	1 NUSM DROSA	P07705 drosophila
13	7	0.9	152	1 PYRI ECOLI	P00478 escherichia
14	7	0.9	152	1 PYRI SALTU	Q82130 salmonella
15	7	0.9	152	1 PYRI SALTU	P08421 salmonella
16	7	0.9	161	1 Y004 BPL2	Q42539 bacterioph
17	7	0.9	200	1 ATP4 IPOBA	Q40089 ipomoea bat
18	7	0.9	212	1 DEOC HALNI	Q9np08 halobacteri
19	7	0.9	225	1 Y116 METJA	Q57580 methanococc
20	7	0.9	230	1 FA3A HUMAN	P98173 homo sapien
21	7	0.9	230	1 FA3A MOUSE	Q948t0 mus musculu
22	7	0.9	231	1 KCY STRCO	Q9eww6 streptomyc
23	7	0.9	249	1 ULA3 HCMVA	P16734 human cytom
24	7	0.9	281	1 YE94_XYLPA	Q9pd85 xyliella fas
25	7	0.9	308	1 HEM4_RHIL0	Q98ei7 rhizobium l
26	7	0.9	315	1 TPIC ARATH	Q9skp6 arabidopsis
27	7	0.9	334	1 PLGJ VIBCH	Q9kg15 vibrio chol
28	7	0.9	335	1 XYNB STRLI	P26515 streptomyc
29	7	0.9	349	1 TRM1 ARCFU	O29443 archaeoglob
30	7	0.9	368	1 MLTA VIBCH	Q9kpq4 vibrio chol
31	7	0.9	383	1 HEM4_PEA	Q43082 pisum sativ
32	7	0.9	388	1 ALR MYCLE	P38056 mycobacteri
33	7	0.9	388	1 PEPF_RABIT	P27823 oryctolagus

Q24093 drosophila
P05783 homo sapien
P34693 caenorhabdi
P33705 canis famli
P47867 mus musculu
P47868 rattus norv
P47455 mycoplasma
Q35280 mus musculu
P53789 oryctolagus
O10439 mouse adeno
P11636 neurospora
P38346 saccharomyc
O67622 aquifex aeo
Q924u5 rattus norv
Q94h13 homo sapien
P07313 oryctolagus
P43634 saccharomyc
P26804 friend mufi
Q17390 caenorhabdi
Q91cc2 anabaena ep
Q92G16 listeria in
Q8Y7q1 listeria mo
Q92i38 rickettsia
P22174 infectious
P53532 corynebacte
Q9es63 mus musculu
Q8k9u3 buchnera ap
Q97ig3 clostridium
Q01454 saccharomyc
Q11168 magnaporthe
P39083 saccharomyc
P55131 actinobacil
Q91783 xenopus lae
Q9vys3 drosophila
Q10435 schizosacch
Q84337 chlamydia t
O43451 homo sapien
P46939 homo sapien
P98161 homo sapien
Q39575 chlamydomon
Q96pk2 homo sapien
P13948 escherichia
P50370 chlamydomon
P39495 bacterioph
P59093 homo sapien
Q78442 guillardia
Q9xsu8 ralatonia s
P03286 human adeno
Q9fnc9 arabidopsis
Q8m9t9 chaetosphe
P81762 pleurotus o
P26538 human papil
P20884 human immun
Q97484 euplotes cr
Q95881 peromyscus
Q95881 peromyscus
Q951f7 arabidopsis
P24957 pisaster oc
P48941 chondrus cr
Q36122 methanobact
Q929H5 thermus the
P14494 escherichia
P14495 escherichia
P14496 escherichia
P10513 escherichia
P04737 escherichia
P01169 lophius ame
P17792 agrobacteri
P43575 saccharomyc
P07480 sus scrofa
Q12370 saccharomyc
P78946 schizosacch
Q11035 mycobacteri
P58349 rhizobium m

107	1	129	6	0.8	1	180	199	1	PSBV_SPIO	1	199	0.8	6	180	P56890 rhizobium m
108	1	131	6	0.8	1	181	201	1	ARP_RAT	1	201	0.8	6	181	Q8t26 methanopyru
109	1	136	6	0.8	1	182	201	1	RECR_UREPA	1	201	0.8	6	182	Q00145 ictaluriid h
110	1	137	6	0.8	1	183	202	1	VNSI_IAPOM	1	202	0.8	6	183	Q29384 archaeoglob
111	1	138	6	0.8	1	184	203	1	Y200_PVRAB	1	203	0.8	6	184	P35775 solenopsis
112	1	141	6	0.8	1	185	203	1	RYLI_VARLI	1	203	0.8	6	185	P45934 bacillus su
113	1	142	6	0.8	1	186	204	1	MDCG_PSEPU	1	204	0.8	6	186	Q9ms69 homo sapien
114	1	143	6	0.8	1	187	204	1	Y275_METUA	1	204	0.8	6	187	Q9kd80 bacillus ha
115	1	146	6	0.8	1	188	208	1	GTP1_CABEL	1	208	0.8	6	188	P07376 salmonella
116	1	146	6	0.8	1	189	208	1	RS6_MYCCE	1	208	0.8	6	189	Q58777 methanococc
117	1	147	6	0.8	1	190	209	1	PCXA_ACICA	1	209	0.8	6	190	P77656 escherichia
118	1	147	6	0.8	1	191	210	1	ALXH_TREPA	1	210	0.8	6	191	P08053 physarum po
119	1	148	6	0.8	1	192	210	1	HIS2_VIBCH	1	210	0.8	6	192	Q60257 methanococc
120	1	148	6	0.8	1	193	211	1	UPP_DEIRA	1	211	0.8	6	193	P43972 haemophilus
121	1	149	6	0.8	1	194	211	1	YD71_MYCPN	1	211	0.8	6	194	Q979u2 clostridium
122	1	153	6	0.8	1	195	212	1	CCMA_HABIN	1	212	0.8	6	195	P05147 emericella
123	1	154	6	0.8	1	196	212	1	KAD_STRAP3	1	212	0.8	6	196	P06973 escherichia
124	1	155	6	0.8	1	197	212	1	KAD_STRP3	1	212	0.8	6	197	P26417 salmonella
125	1	155	6	0.8	1	198	212	1	KAD_STRP3	1	212	0.8	6	198	Q08081 meriones un
126	1	155	6	0.8	1	199	212	1	KAD_STRP3	1	212	0.8	6	199	P04868 barley stri
127	1	155	6	0.8	1	200	212	1	KAD_STRP3	1	212	0.8	6	200	Q8xdk1 escherichia
128	1	155	6	0.8	1	201	212	1	Y434_RICPR	1	212	0.8	6	201	Q8fjy3 escherichia
129	1	155	6	0.8	1	202	212	1	YGBL_ECOLI	1	212	0.8	6	202	P77385 escherichia
130	1	157	6	0.8	1	203	212	1	YJ47_ARCFU	1	212	0.8	6	203	Q9xhm3 corynebacte
131	1	158	6	0.8	1	204	213	1	MDCG_XANAC	1	213	0.8	6	204	P18317 klebsiella
132	1	159	6	0.8	1	205	214	1	EPD_NOTCH	1	214	0.8	6	205	Q8x7n1 escherichia
133	1	159	6	0.8	1	206	215	1	CYB5_CVACA	1	215	0.8	6	206	P32056 escherichia
134	1	159	6	0.8	1	207	215	1	EPD1_CARAU	1	215	0.8	6	207	Q54673 rhodobacter
135	1	160	6	0.8	1	208	215	1	EPD2_CARAU	1	215	0.8	6	208	Q84842 paprika ml
136	1	160	6	0.8	1	209	215	1	EPD_CYPCA	1	215	0.8	6	209	Q83486 tobamovirus
137	1	160	6	0.8	1	210	215	1	ERD2_ARATH	1	215	0.8	6	210	P71685 mycobacteri
138	1	160	6	0.8	1	211	215	1	ERD2_PETHY	1	215	0.8	6	211	Q8diko synecococc
139	1	161	6	0.8	1	212	215	1	ORN_MYCLE	1	215	0.8	6	212	P06450 mycobacteri
140	1	161	6	0.8	1	213	215	1	ORN_MYCTU	1	215	0.8	6	213	P31451 escherichia
141	1	161	6	0.8	1	214	216	1	YGM8_YEAST	1	216	0.8	6	214	P4275 haemophilus
142	1	161	6	0.8	1	215	217	1	DEOC_MYCHO	1	217	0.8	6	215	Q8xwc5 ralstonia s
143	1	161	6	0.8	1	216	217	1	EPD_BRARE	1	217	0.8	6	216	P22911 oryza sativ
144	1	164	6	0.8	1	217	217	1	FDNI_ECOLI	1	217	0.8	6	217	P22912 oryza sativ
145	1	164	6	0.8	1	218	217	1	VNS1_IANAN	1	217	0.8	6	218	P73527 synecocyst
146	1	164	6	0.8	1	219	218	1	VNS1_IACKO	1	218	0.8	6	219	P39789 bacillus su
147	1	165	6	0.8	1	220	218	1	EPD_DANAE	1	218	0.8	6	220	Q08691 staphylococ
148	1	165	6	0.8	1	221	218	1	SGAR_MYCPN	1	218	0.8	6	221	Q08021 salmonella
149	1	168	6	0.8	1	222	218	1	TRPF_BACHD	1	218	0.8	6	222	P50829 bacillus su
150	1	170	6	0.8	1	223	218	1	Y4VH_RHISN	1	218	0.8	6	223	Q9pev2 xyella fas
151	1	171	6	0.8	1	224	220	1	FGF3_CHICK	1	220	0.8	6	224	Q9y7j4 schizosacch
152	1	174	6	0.8	1	225	221	1	YGM1_HABIN	1	221	0.8	6	225	Q8u9r3 brachydanio
153	1	174	6	0.8	1	226	222	1	PLA5_PVRAB	1	222	0.8	6	226	P13421 serratia ma
154	1	174	6	0.8	1	227	222	1	NANE_STAMM	1	222	0.8	6	227	P66400 aquifex ase
155	1	175	6	0.8	1	228	222	1	NANE_STAMM	1	222	0.8	6	228	P21888 pantoea ana
156	1	176	6	0.8	1	229	226	1	CG5_MOUSE	1	226	0.8	6	229	Q67587 aquifex ase
157	1	178	6	0.8	1	230	227	1	TPIS_SULSO	1	227	0.8	6	230	Q24251 drosophila
158	1	178	6	0.8	1	231	227	1	VNS1_IACKJ	1	227	0.8	6	231	Q28533 archaeoglob
159	1	178	6	0.8	1	232	227	1	VNS1_IADU3	1	227	0.8	6	232	Q8xjb2 clostridium
160	1	179	6	0.8	1	233	227	1	VNS1_IAMU3	1	227	0.8	6	233	Q8yvh1 anabaena sp
161	1	180	6	0.8	1	234	227	1	VNS1_IATRS	1	227	0.8	6	234	P76471 escherichia
162	1	180	6	0.8	1	235	227	1	VNS1_IATRS	1	227	0.8	6	235	P49694 thermomonas
163	1	182	6	0.8	1	236	228	1	TRMD_UREPA	1	228	0.8	6	236	Q9p06 campylobact
164	1	183	6	0.8	1	237	230	1	VNS1_CVHM	1	230	0.8	6	237	Q9f414 staphylococ
165	1	183	6	0.8	1	238	230	1	VNS1_CVHOC	1	230	0.8	6	238	Q8vr31 methylococ
166	1	183	6	0.8	1	239	230	1	VNS1_CVHOC	1	230	0.8	6	239	P03739 bacterioph
167	1	183	6	0.8	1	240	230	1	VNS1_IACKB	1	230	0.8	6	240	P56312 chlorella v
168	1	184	6	0.8	1	241	230	1	VNS1_IACKB	1	230	0.8	6	241	P87068 laccaria bi
169	1	188	6	0.8	1	242	230	1	VNS1_IACKB	1	230	0.8	6	242	Q9hrt1 halobacteri
170	1	190	6	0.8	1	243	230	1	VNS1_IACKB	1	230	0.8	6	243	P73869 synecocyst
171	1	190	6	0.8	1	244	230	1	VNS1_IACKB	1	230	0.8	6	244	Q53021 erwinia chr
172	1	193	6	0.8	1	245	230	1	VNS1_IACKB	1	230	0.8	6	245	Q9kqh1 vibrio chol
173	1	196	6	0.8	1	246	230	1	VNS1_IACKB	1	230	0.8	6	246	Q8xau9 escherichia
174	1	196	6	0.8	1	247	230	1	VNS1_IACKB	1	230	0.8	6	247	P75894 escherichia
175	1	197	6	0.8	1	248	230	1	VNS1_IACKB	1	230	0.8	6	248	Q9pan4 xyella fas
176	1	197	6	0.8	1	249	231	1	VNS1_IACKB	1	231	0.8	6	249	Q9ju99 neisseria m
177	1	197	6	0.8	1	250	232	1	VNS1_IACKB	1	232	0.8	6	250	Q9jz78 neisseria m
178	1	197	6	0.8	1	251	232	1	VNS1_IACKB	1	232	0.8	6	251	P35651 ralstonia s
179	1	198	6	0.8	1	252	232	1	VNS1_IACKB	1	232	0.8	6	252	Q28374 archaeoglob

253	6	0.8	233	1	MENH_BACSU	P31113	326	1	HSLO_STAAM	Q99w91	staphylococ
254	6	0.8	236	1	ABME_RABIT	P47855	327	6	HSLO_STAAM	Q8nx23	staphylococ
255	6	0.8	237	1	LIVF_ECOLI	P22731	328	6	YSEI_ECOLI	P32484	escherichia
256	6	0.8	237	1	VNSI_IAAIC	Q02599	329	6	DAPA_THEME	Q8x1k9	thermotoga
257	6	0.8	237	1	VNSI_IACHI	P11618	330	6	MALK_ENTAE	P18813	enterobacte
258	6	0.8	237	1	VNSI_IAPOF	P03497	331	6	RRK2_FORPU	P51311	porphyra pu
259	6	0.8	237	1	VNSI_IAUSS	P03498	332	6	DAPA_OCEIH	Q8eg11	oceanobacil
260	6	0.8	238	1	RL2_AERPE	Q9vfn1	333	6	HTPX_LEPIN	Q8exn4	leptosira
261	6	0.8	238	1	YJHA_ECOLI	P39372	334	6	ISPE_XANCP	Q8pc64	xanthomonas
262	6	0.8	240	1	VP26_NPVAC	P08358	335	6	MOAA_ARCFU	Q28273	archaeoglob
263	6	0.8	244	1	APMI_HUMAN	Q15848	336	6	PCOB_ECOLI	Q47453	escherichia
264	6	0.8	244	1	OPAI_NEIGO	Q04877	337	6	TPIC_SECEE	P46223	secale cere
265	6	0.8	246	1	YIAT_ECO57	P58224	338	6	YIHV_ECOLI	P32143	escherichia
266	6	0.8	246	1	YIAT_ECOLI	P37681	339	6	HTPX_STRPN	Q79td6	streptococc
267	6	0.8	247	1	APMI_MOUSE	Q60994	340	6	HTPX_STRR6	Q8dph5	streptococc
268	6	0.8	249	1	KDKA_XANAC	Q8phb3	341	6	RIP3_MAIZE	P25891	zea mays (m
269	6	0.8	249	1	TL3X_MOUSE	Q9et35	342	6	FLIP_RHOSH	Q85133	rhodobacter
270	6	0.8	251	1	HEMA_PSEAE	P48246	343	6	RIPX_MAIZE	P28522	zea mays (m
271	6	0.8	252	1	PRGK_SALTY	P41786	344	6	FDHE_SHEON	P59188	shewanella
272	6	0.8	252	1	PYHD_CVPOP	P36701	345	6	YET6_YEAR	P40065	saccharomyc
273	6	0.8	253	1	ATP6_TRIRU	Q36835	346	6	CHEV_BACSU	P37599	bacillus su
274	6	0.8	255	1	HA21_MOUSE	P01904	347	6	RIP9_MAIZE	P25892	zea mays (m
275	6	0.8	255	1	HA22_MOUSE	P04224	348	6	Y007_MYCTU	P71575	mycobacteri
276	6	0.8	255	1	RS2_THETH	P80371	349	6	NAC_ECOLI	Q47005	escherichia
277	6	0.8	255	1	Y166_METJA	Q57630	350	6	NAC_KLEAE	Q85997	klebsiella
278	6	0.8	256	1	ATP6_HANWI	P48879	351	6	VP53_BPAPS	Q9tip5	bacterioph
279	6	0.8	256	1	GRST_BACBR	P14686	352	6	FLGJ_VIBPA	Q9x9j3	vibrio para
280	6	0.8	256	1	SURE_AGR5	Q8ueg3	353	6	MEAW_UREPA	Q9pqa4	ureaplasma
281	6	0.8	258	1	RSEA_BACSU	P39650	354	6	Y117_HELPJ	Q9zmv6	helicobacte
282	6	0.8	260	1	RFBA_MYXKA	Q50862	355	6	Y117_HELPY	P56080	helicobacte
283	6	0.8	261	1	Y278_PSEAE	P03256	356	6	NTCB_SYNY3	P74422	synecocyst
284	6	0.8	261	1	YZ78_PSEAE	Q9hy42	357	6	YJ09_AQUAE	Q67744	aquifex aeo
285	6	0.8	262	1	Y4JP_RHISN	P55516	358	6	YX48_STRCO	Q9x8h1	streptomyce
286	6	0.8	265	1	DH47_ATHAT	P31168	359	6	RFAD_ECOLI	P17963	escherichia
287	6	0.8	266	1	NAGB_VIBCH	Q9kks5	360	6	ESCA_ATHAT	Q987c9	arabidopsis
288	6	0.8	266	1	NAGB_VIBVU	Q8d4t9	361	6	CELF_PRVKA	Q85230	pseudorabie
289	6	0.8	266	1	THPS_LACCA	P17166	362	6	TRXB_CHLTR	Q41011	chlamydia t
290	6	0.8	267	1	DHPS_STAAR	Q05701	363	6	VR2B_BPT4	P33691	bacterioph
291	6	0.8	267	1	NADP_PYRAE	Q8zx14	364	6	Y353_BUCAP	Q8k912	buchnera ap
292	6	0.8	267	1	RL4_SULSO	Q9uxa6	365	6	CBR1_RHIME	P58332	rhizobium m
293	6	0.8	267	1	SV72_ATHAT	Q94kk6	366	6	CBR2_RHIME	P56885	rhizobium m
294	6	0.8	268	1	HLVC_TREHY	Q54318	367	6	PLDB_HABTN	P44800	haemophilus
295	6	0.8	268	1	IL1B_AERPE	P14628	368	6	CSBT_SULSO	P58030	sulfolobus
296	6	0.8	268	1	SUHB_AERPE	Q9yaz7	369	6	TPIC_FRAAN	Q9m4s8	fragaria an
297	6	0.8	269	1	DAPF_CHLNU	Q9pjw2	370	6	LDH_CLOPE	Q8xp62	clostridium
298	6	0.8	269	1	PYHD_CPVES	P36326	371	6	EX53_MYCTU	Q10699	mycobacteri
299	6	0.8	270	1	URED_KLEAE	Q09063	372	6	FTBP_ADE03	P04501	human adeno
300	6	0.8	270	1	URED_KLEPN	Q02944	373	6	PE13_ATHAT	Q49293	arabidopsis
301	6	0.8	270	1	YNF0_YEAST	Q02944	374	6	RCCR_ATHAT	Q81du4	arabidopsis
302	6	0.8	273	1	OSAE_BORBU	P53952	375	6	HEMZ_XANAC	Q8pex0	xanthomonas
303	6	0.8	274	1	COX1_CHOBI	P50668	376	6	MPC2_ALCEU	P17296	alcaligenes
304	6	0.8	274	1	COX1_CHOFU	P50689	377	6	CCSA_MAIZE	P46659	zea mays (m
305	6	0.8	274	1	COX1_CHOOC	P50670	378	6	CCSA_ORYSA	P12215	oryza sariv
306	6	0.8	274	1	COX1_CHORO	P50671	379	6	CCSA_WHEAT	P58266	tritium ae
307	6	0.8	274	1	COX3_ALLMA	P80439	380	6	CCSA_LOTJA	Q9bbp4	lotus japon
308	6	0.8	275	1	CTA2_ACLIW	Q33950	381	6	CCSA_SPIOL	Q9m3j1	spinacia ol
309	6	0.8	275	1	FABI_HELPJ	Q9zmn7	382	6	COBD_PSEDE	P21634	pseudomonas
310	6	0.8	276	1	BSP4_CABEL	P46759	383	6	IN30_PEA	Q03943	pisum sativ
311	6	0.8	276	1	YZ36_AQUAE	Q87427	384	6	CCSI_CAEEL	P12114	caenorhabdi
312	6	0.8	277	1	RSU1_HUMAN	Q15404	385	6	LEGA_RAT	P38552	rattus norv
313	6	0.8	277	1	RSU1_MOUSE	Q01730	386	6	YXDK_BACSU	P42422	bacillus su
314	6	0.8	278	1	P29K_STRGC	P42361	387	6	YGRG_ECOLI	P52052	escherichia
315	6	0.8	281	1	RL2_MYCCA	P10133	388	6	ASCD_YERPE	P37911	versinia pe
316	6	0.8	282	1	YC53_METJA	Q58650	389	6	CCSA_ATHAT	P56770	arabidopsis
317	6	0.8	283	1	YB89_METJA	Q58589	390	6	YEIR_ECOLI	P33030	escherichia
318	6	0.8	284	1	KPRS_PYRAE	Q8zu24	391	6	RRM1_HUMAN	Q9uet6	homo sapien
319	6	0.8	284	1	RS2_MYCCE	P47316	392	6	HEBP_ATHAT	P43273	arabidopsis
320	6	0.8	286	1	TRT2_SHEEP	P50751	393	6	SAPD_SALTY	P36636	salmonella
321	6	0.8	288	1	OTX2_XENLA	Q91813	394	6	LDHA_CHICK	P00340	gallus gall
322	6	0.8	289	1	DAPA_METJA	Q57695	395	6	MDHC_ATHAT	P93819	arabidopsis
323	6	0.8	291	1	HSLO_BACHD	Q3kgh4	396	6	MDHC_BETVU	Q9sm18	beta vulgar
324	6	0.8	291	1	HSLO_BACSU	P37565	397	6	MDHC_MAIZE	Q08062	zea mays (m
325	6	0.8	291	1	LIP_THELA	O59952	398	6	MDHC_MEDSA	O48905	medicago sa

399	6	0.8	332	1	MDHC_MESSR	Q24047	mesembryant	472	6	0.8	378	1	FLAB_VIBPA	Q56702	vibrio par
400	6	0.8	332	1	MDHD_ARATH	Q24047	arabidopsis	473	6	0.8	378	1	FLAE_VIBCH	Q6kq60	vibrio choi
401	6	0.8	333	1	DCUD_THEAC	Q57106	thermoplasm	474	6	0.8	379	1	FLAA_VIBAN	Q60246	vibrio angu
402	6	0.8	333	1	TRPD_PYRAE	Q9hb99	pyrobaculum	475	6	0.8	379	1	FLAA_VIBCH	Q30858	vibrio choi
403	6	0.8	334	1	Y378_METTH	Q8zw45	methanobact	476	6	0.8	379	1	IDHB_CAEEL	Q33353	caenorhabdi
404	6	0.8	335	1	GCAT_ABRHY	P10480	aeromonas h	477	6	0.8	379	1	P2X6_MOUSE	Q54803	mus muscullu
405	6	0.8	335	1	MRAY_CHLMU	Q9plg6	chlamydia m	478	6	0.8	379	1	P2X6_MOUSE	P51579	rattus norv
406	6	0.8	337	1	LPXD_XANAC	Q8pm15	xanthomonas	479	6	0.8	380	1	APG2_ARCFU	Q28847	archaeoglob
407	6	0.8	337	1	LPXD_XANCP	Q8paw3	xanthomonas	480	6	0.8	380	1	CAH2_CHLRE	Q24258	chlamydomon
408	6	0.8	337	1	Y020_BUCAP	Q44634	buchnera ap	481	6	0.8	380	1	SUC2_ARCFU	P28097	archaeoglob
409	6	0.8	338	1	ANK9_MOUSE	Q9fjhg0	mus muscullu	482	6	0.8	380	1	YK84_MYCTU	O10692	mycobacteri
410	6	0.8	338	1	HMDX_METJA	Q58125	methanococc	483	6	0.8	381	1	CPXG_STRSQ	P23296	streptomyc
411	6	0.8	338	1	SDHL_YEAST	P17324	saccharomyc	484	6	0.8	381	1	CYB_CAEFU	Q9zy46	caenolestes
412	6	0.8	338	1	SYW_VIBCH	Q9kv7	vibrio chol	485	6	0.8	381	1	YBD2_YEAST	P38199	saccharomyc
413	6	0.8	338	1	SYW_VIBPA	Q87113	vibrio para	486	6	0.8	382	1	CYB_DIDAL	Q34279	didelphis a
414	6	0.8	338	1	SYW_VIBVU	Q8dct6	vibrio vuln	487	6	0.8	382	1	FATB_UMBCA	Q41635	umbelluluri
415	6	0.8	338	1	YJEK_HAEIN	P44641	haemophilus	488	6	0.8	382	1	Y022_NPVOP	O10281	orgyia pseu
416	6	0.8	340	1	ASPG_FLAME	Q47898	flavobacter	489	6	0.8	382	1	YSVF_CAEEL	Q8ta81	caenorhabdi
417	6	0.8	340	1	BYR1_SCHPO	P10506	schizosacch	490	6	0.8	382	1	SUC2_STRCO	Q87840	streptomyc
418	6	0.8	342	1	Y159_TREPA	Q83194	treponema p	491	6	0.8	383	1	DDL_MYCLE	O9cbs0	mycobacteri
419	6	0.8	342	1	YJEK_ECOLI	P39280	escherichia	492	6	0.8	384	1	SUC1_STROCO	Q925h8	mycobacteri
420	6	0.8	343	1	HMD_METVO	Q50840	methanococc	493	6	0.8	387	1	NMT_DROME	P57245	buchnera ap
421	6	0.8	345	1	ARGC_BACHD	Q9kv82	bacillus ha	494	6	0.8	387	1	SUC2_MYCTU	O61613	drosofila
422	6	0.8	345	1	YFIF_ECOLI	P33635	escherichia	495	6	0.8	387	1	TI22_MOUSE	P71559	mycobacteri
423	6	0.8	347	1	GALE_RAT	P18645	rattus norv	496	6	0.8	388	1	COAT_AMCV	Q9eqn3	mus muscullu
424	6	0.8	347	1	PUR7_HALN1	Q9hnu7	halobacteri	497	6	0.8	388	1	DUT_HCMVA	P14836	artichoke m
425	6	0.8	348	1	PEN_METKA	Q8txu4	methanopyru	498	6	0.8	389	1	RFIM_CAEEL	P16824	human cytom
426	6	0.8	348	1	GALE_HUMAN	Q14376	homo sapien	499	6	0.8	390	1	GASA_XENLA	O44568	caenorhabdi
427	6	0.8	349	1	RFCE_MOUSE	Q9wuk4	mus muscullu	500	6	0.8	391	1	ACKA_BACHD	P43695	xenopus lae
428	6	0.8	350	1	DDL_WIGBR	Q8dzv5	wiggleswort	501	6	0.8	392	1	PNCB_YNAC	O9k815	bacillus ha
429	6	0.8	351	1	OM32_CMAC	P24305	comamonas a								

545	6	0.8	417	1	YCHO_ECOLI	P39165	escherichia	618	1	G6PI_STRP8	Q9p2r3	streptococ
546	6	0.8	418	1	PROA_CLOAB	Q97e62	clostridium	619	1	G6PI_STRPY	Q9a1l1	streptococ
547	6	0.8	419	1	ENO_PYRAE	Q87ye7	pyrobaculum	620	1	HEMN_BRAJA	Q31381	bradyrhizo
548	6	0.8	419	1	VEGC_HUMAN	P49767	homo sapien	621	1	UTH1_YEAST	Q36135	saccharomyc
549	6	0.8	420	1	PUR2_STRPN	Q97f98	streptococ	622	1	RATM_YEAST	Q01802	saccharomyc
550	6	0.8	420	1	PUR2_STRSU	Q97f98	streptococ	623	1	PPOX_MYCLE	Q50008	mycobacteri
551	6	0.8	420	1	SYW_ARCFU	O28579	archaeoglob	624	1	HOS2_YEAST	P53096	saccharomyc
552	6	0.8	421	1	PUR2_STRP3	O88y74	streptococ	625	1	ENGA_ANASP	Q8y2h7	anaena sp
553	6	0.8	421	1	PUR2_STRP8	O8p308	streptococ	626	1	SSF1_YEAST	P38789	saccharomyc
554	6	0.8	421	1	PUR2_STRPY	O8aiy7	streptococ	627	1	SSF2_YEAST	Q12153	saccharomyc
555	6	0.8	421	1	Y118_TREPA	O83155	treponema p	628	1	TNA1_SYMTH	P31014	symbiobacte
556	6	0.8	422	1	HEAD_BPAPS	Q9t1s4	bacterioph	629	1	GLNA_METJA	Q60182	methanococ
557	6	0.8	422	1	PUR2_BACSU	P12039	bacillus su	630	1	GUAD_HUMAN	Q9y2t3	homo sapien
558	6	0.8	423	1	PUR2_NEIMA	O9jw6	neisseria m	631	1	PHR_STRGR	P12768	streptomyce
559	6	0.8	423	1	PUR2_NEIMA	O9jxa3	neisseria m	632	1	SIN_MYCPN	P75521	mycoplasma
560	6	0.8	424	1	CUBO_CAEEL	O8w51	caenorhabdi	633	1	YAO6_SCHPO	Q10085	schizosacch
561	6	0.8	424	1	THIL_RAT	P17764	rattus norv	634	1	PROA_YEAST	P54885	saccharomyc
562	6	0.8	425	1	POXN_DROME	P23758	drosophila	635	1	REP_STRLI	P22406	streptomyce
563	6	0.8	425	1	TANK_HUMAN	Q92844	homo sapien	636	1	CD4_SAISC	Q23037	saimiri sci
564	6	0.8	426	1	SLS1_YARLI	Q99158	yarrowia li	637	1	MURD_DEIRA	Q9rrt4	deinococcus
565	6	0.8	426	1	YMW8_YEAST	Q04749	saccharomyc	638	1	DB80_DROME	O61305	drosophila
566	6	0.8	427	1	CYB6_CHLIT	O59297	chlorobium	639	1	DVR_SCHPO	P36591	schizosacch
567	6	0.8	427	1	CYB6_CHLTE	Q9f721	chlorobium	640	1	PEX2_PICPA	Q01964	pichia past
568	6	0.8	427	1	GATD_AERPE	O9yrt8	aeropyrum p	641	1	DALD_YERPE	P58709	versinia pe
569	6	0.8	427	1	TCO2_HUMAN	P20062	homo sapien	642	1	GALP_ECOLI	P37021	escherichia
570	6	0.8	427	1	TCO2_RAT	Q9r0d6	rattus norv	643	1	Y065_MYCPN	P75612	mycoplasma
571	6	0.8	427	1	THIL_HUMAN	P24752	homo sapien	644	1	DLD3_PSRPU	P31046	pseudomonas
572	6	0.8	428	1	HEMA_CVMA5	P31615	murine coro	645	1	DNAA_PROMI	P22837	proteus mir
573	6	0.8	430	1	TCO2_MOUSE	O88968	mus musculu	646	1	Y065_MYCGE	P47311	mycoplasma
574	6	0.8	431	1	AROQ_AQUAE	O67494	aquifex aeo	647	1	TM11_MOUSE	Q99p22	mus musculu
575	6	0.8	431	1	P2X6_HUMAN	O15547	homo sapien	648	1	SG3_HUMAN	Q8wx22	homo sapien
576	6	0.8	431	1	PUR8_BACSU	P12047	bacillus su	649	1	ARI1_MOUSE	Q921k5	mus musculu
577	6	0.8	431	1	SYW_XANCP	Q8p3z4	xanthomonas	650	1	SECY_SULSO	Q9ux84	sulfolobus
578	6	0.8	431	1	TIG_CLOAB	Q97ft6	clostridium	651	1	YOJ8_CAEEL	P34631	caenorhabdi
579	6	0.8	431	1	UL78_HCNVA	P16751	human cytom	652	1	HQCT_RAUSE	Q9ar73	rauvolfia s
580	6	0.8	432	1	TCO2_BOVIN	Q9xsc9	bos taurus	653	1	CP71_PEARA	P24465	persea amer
581	6	0.8	433	1	SYD_HALVO	O24822	halobacteri	654	1	GBA1_YEAST	P08539	saccharomyc
582	6	0.8	435	1	AKR_ARATH	Q05753	arabidopsis	655	1	FXH1_BRARE	Q919el	brachydanio
583	6	0.8	435	1	DHOM_METGL	P37143	methylobaci	656	1	GLGS_WHEAT	P35338	saccharomyc
584	6	0.8	435	1	FUZ7_USTWA	Q99078	ustilago ma	657	1	MAIR_YEAST	P10508	saccharomyc
585	6	0.8	436	1	SH6_RAT	P31388	rattus norv	658	1	MA6R_YEAST	P08401	escherichia
586	6	0.8	436	1	ENV_FLVCS	Q02077	feline leuk	659	1	CREC_ECOLI	P47870	homo sapien
587	6	0.8	436	1	BRFA_ARATH	Q39097	arabidopsis	660	1	GAB2_HUMAN	P15432	mus musculu
588	6	0.8	437	1	AP2A_HUMAN	P05549	homo sapien	661	1	GAB2_MOUSE	P05053	escherichia
589	6	0.8	437	1	AP2A_MOUSE	P34056	mus musculu	662	1	KLF4_MOUSE	Q60793	mus musculu
590	6	0.8	437	1	AP2A_RAT	P58197	rattus norv	663	1	PTGB_ECOLI	P05053	escherichia
591	6	0.8	437	1	ERF1_HUMAN	P46055	homo sapien	664	1	PYR5_DICDI	P09556	dictyosteli
592	6	0.8	437	1	ERF1_XENLA	P35615	xenopus lae	665	1	S61A_SCHPO	P79088	schizosacch
593	6	0.8	437	1	SECY_STRCO	P46785	streptomyce	666	1	PTMB_CLOAB	O65989	clostridium
594	6	0.8	437	1	SECY_STRGB	Q59912	streptomyce	667	1	DUSA_HUMAN	Q9Y5W6	homo sapien
595	6	0.8	437	1	SECY_STRGR	Q59916	streptomyce	668	1	DUSA_MOUSE	Q9580	mus musculu
596	6	0.8	437	1	SECY_STRLI	P49977	streptomyce	669	1	GLGS_ORISA	P15280	oryza sativ
597	6	0.8	437	1	SECY_STRSC	P43416	streptomyce	670	1	SYE_DEIRA	Q9Y5W6	homo sapien
598	6	0.8	438	1	GATD_PYRAB	Q9V0T9	pyrococcus	671	1	NCAP_HAZVU	P27318	hazara viru
599	6	0.8	438	1	GATD_PYRHO	O59132	pyrococcus	672	1	BCHZ_RHOGE	Q9Y5W6	homo sapien
600	6	0.8	439	1	EX7L_HAEIN	P43913	haemophilus	673	1	MDM2_CANFA	Q9Y5W6	homo sapien
601	6	0.8	440	1	SH6_HUMAN	P50406	homo sapien	674	1	GALT_LACHE	P06950	canis fami
602	6	0.8	440	1	SH6_MOUSE	Q9rlc8	mus musculu	675	1	YK11_CAEEL	Q00054	lactobacill
603	6	0.8	440	1	MURD_VIECH	Q9kpg5	vibrio chol	676	1	DLAB_ECOLI	P34312	caenorhabdi
604	6	0.8	441	1	TOXE_COCCA	O74205	cochliobolu	677	1	GLGS_BETVU	P17445	escherichia
605	6	0.8	443	1	GLNA_CLOSA	P10656	clostridium	678	1	ET83_DROME	P55232	beta vulgar
606	6	0.8	444	1	GUNN_ERWCA	Q59394	erwinia car	679	1	AP2A_SHEEP	P29774	drosophila
607	6	0.8	445	1	DNAA_ACHLA	Q9khu8	acholeplasm	680	1	TYTR_TRYCR	Q9N0n3	ovis aries
608	6	0.8	445	1	MRSA_HAEIN	P54164	haemophilus	681	1	ANL2_MOUSE	P28593	trypanosoma
609	6	0.8	446	1	GLNA_METVO	P21154	methanococc	682	1	ANYA_DROMA	Q9R045	mus musculu
610	6	0.8	446	1	PIV2_ADECC	Q55945	canine aden	683	1	ANYA_DROME	P54215	drosophila
611	6	0.8	446	1	PIV2_ADECT	P87552	canine aden	684	1	ANYB_DROME	P81641	drosophila
612	6	0.8	447	1	CLUS_RAT	O53371	rattus norv	685	1	TEBA_OXYNO	P29549	oxytricha n
613	6	0.8	447	1	DTA3_RALSO	Q8xr66	ralstonia s	686	1	LYXK_ECOLI	P37577	escherichia
614	6	0.8	447	1	HWC8_SCHPO	P54874	schizosacch	687	1	OPSI_PATYE	Q15973	patinopecte
615	6	0.8	448	1	CLUS_MOUSE	Q06890	mus musculu	688	1	DP22_RAT	Q96b1	rattus norv
616	6	0.8	448	1	HEM1_METWA	Q8pw60	methanosarc	689	1	SECD_AQUAE	O67102	aquifex aeo
617	6	0.8	449	1	G6PI_STRP3	Q8k8q6	streptococc	690	1	PIV2_ADECR	Q96680	canine aden

691	6	0.8	503	1	CATA_MICLU	P29422 micrococ	764	0.8	566	1	CC45_HUMAN	O75419 homo sapien
692	6	0.8	503	1	SECD_HELPY	O26074 helicobacte	765	0.8	566	1	CC45_MOUSE	O821x9 mus musculus
693	6	0.8	503	1	SNAL_MOUSE	Q61234 mus musculus	766	0.8	569	1	ATKB_SALTI	O828e5 salmonella
694	6	0.8	504	1	GUNW_ERWCA	O59395 erwinia car	767	0.8	569	1	HYCE_ECOLI	P16431 escherichia
695	6	0.8	505	1	GUNW_ERWCA	O47096 erwinia car	768	0.8	572	1	GPC5_HUMAN	P78333 homo sapien
696	6	0.8	505	1	SNAL_HUMAN	Q13424 homo sapien	769	0.8	575	1	ESR2_ICTPU	Q91aki ictalurus p
697	6	0.8	505	1	SNAL_RABIT	Q28626 oryctolagus	770	0.8	576	1	ACEA_RICCO	P15479 ricinus com
698	6	0.8	505	1	YOD2_SCHPO	Q9c0v0 schizosacch	771	0.8	576	1	ACH2_DROME	P17644 drosophila
699	6	0.8	506	1	DLTA_LACRH	P35854 lactobacill	772	0.8	576	1	GRK6_HUMAN	P43250 homo sapien
700	6	0.8	506	1	HITB_HABIN	P71338 haemophilus	773	0.8	576	1	GRK6_MOUSE	O70293 mus musculus
701	6	0.8	507	1	CHLB_PORPU	P51278 porphyra pu	774	0.8	576	1	GRK6_RAT	P97711 rattus norv
702	6	0.8	508	1	CC37_CANAL	Q8x1e6 candida alb	775	0.8	576	1	RMIN_DROME	P95527 drosophila
703	6	0.8	508	1	GLGS_VICPA	P24166 vicia faba	776	0.8	577	1	SYR_SALTI	Q825v7 salmonella
704	6	0.8	509	1	C931_SOYBN	Q42798 glycine max	777	0.8	577	1	SYR_SALTY	P74871 salmonella
705	6	0.8	510	1	C933_SOYBN	O81973 glycine max	778	0.8	577	1	YG5U_YEAST	P53333 saccharomyc
706	6	0.8	511	1	EGO_ECOLI	P77257 escherichia	779	0.8	579	1	YCSJ_BACSU	P42967 bacillus su
707	6	0.8	512	1	G63A_DROME	Q9vz17 drosophila	780	0.8	579	1	PHIA_ECOLI	Q47153 escherichia
708	6	0.8	512	1	GLGT_VICFA	P52417 vicia faba	781	0.8	579	1	RMDH_MAIZE	O24594 zea mays (m
709	6	0.8	513	1	GLGS_HORVU	P25238 hordeum vul	782	0.8	580	1	ACEA_PINTA	Q43097 pinus taeda
710	6	0.8	513	1	SYS_METH	O27194 methanobact	783	0.8	584	1	ACES_RABIT	Q29499 oryctolagus
711	6	0.8	514	1	CBH_SCHPO	O14423 schizosacch	784	0.8	586	1	CO9_FUGRU	P79755 fugu rubrip
712	6	0.8	514	1	CTP7_ICTPU	Q73853 ictalurus p	785	0.8	588	1	CALI_BOVIN	Q28068 bos taurus
713	6	0.8	514	1	IMA3_CABEL	Q19969 caenorhabdi	786	0.8	588	1	CALI_HUMAN	Q13939 homo sapien
714	6	0.8	515	1	LEU1_HABIN	P43861 haemophilus	787	0.8	588	1	DECA_DROME	P07713 drosophila
715	6	0.8	515	1	YWCA_BACSU	P29599 bacillus su	788	0.8	590	1	GRK5_HUMAN	P34947 homo sapien
716	6	0.8	516	1	LADI_HUMAN	O00515 homo sapien	789	0.8	590	1	GRK5_RAT	Q62833 rattus norv
717	6	0.8	518	1	LEU1_PASMO	Q9cjm5 pasteurella	790	0.8	590	1	SYD_SALTY	Q825w1 salmonella
718	6	0.8	518	1	SAP_CHICK	O13035 gallus gall	791	0.8	590	1	SYD_SALTY	Q825w2 salmonella
719	6	0.8	518	1	TEX5_HUMAN	Q99593 homo sapien	792	0.8	591	1	SYD_PSEAR	Q51422 pseudomonas
720	6	0.8	518	1	TEX5_MOUSE	P70326 mus musculus	793	0.8	591	1	SYD_VIECH	Q9ksu0 vibrio chol
721	6	0.8	520	1	CZCB_ALCEU	P13510 alcaligenes	794	0.8	592	1	SYD_VIBPA	Q87qw2 vibrio para
722	6	0.8	520	1	CZCB_ALCEU	P24176 alcaligenes	795	0.8	592	1	SYD_VIBPA	Q87qw2 vibrio para
723	6	0.8	520	1	GLGS_ARATH	P52228 arabidopsis	796	0.8	594	1	NUSM_HIPAM	Q92zy1 hippopotamu
724	6	0.8	520	1	GLGS_BRANA	Q9m462 brassica na	797	0.8	595	1	BETP_CORGL	P54582 corynebacte
725	6	0.8	521	1	GLGS_LYCES	Q42882 lycopersico	798	0.8	595	1	BPRX_BACNO	P24280 bacteroides
726	6	0.8	521	1	GLGS_SOLTU	P23509 solanum tub	799	0.8	595	1	DCOA_KLEPN	P13187 klebsiella
727	6	0.8	525	1	GOR_PANTR	P48778 pan troglod	800	0.8	600	1	ESRI_RAT	P06211 rattus norv
728	6	0.8	526	1	SECD_HELPY	Q9zj66 helicobacte	801	0.8	602	1	YZ13_METJA	Q60275 methanococc
729	6	0.8	527	1	NPRE_BACBR	P32633 bacillus br	802	0.8	602	1	PURL_THEMA	Q9x0x3 thermotoga
730	6	0.8	528	1	YRP3_GIALA	P25203 giardia lam	803	0.8	604	1	YJ13_YEAST	P47030 saccharomyc
731	6	0.8	529	1	SWA_DROPS	Q9u915 drosophila	804	0.8	605	1	PESC_YEAST	P53261 saccharomyc
732	6	0.8	532	1	CRTI_APHSP	P21134 aphanocep	805	0.8	606	1	CACI_YEAST	Q12495 saccharomyc
733	6	0.8	536	1	CATA_DEIRA	Q59337 deinococcus	806	0.8	607	1	HEMA_PHODV	P28862 phocine dis
734	6	0.8	537	1	CEK1_HUMAN	Q8ctc0 homo sapien	807	0.8	607	1	XYNA_NEOPA	P29127 neocallimas
735	6	0.8	538	1	CK11_YEAST	P3291 saccharomyc	808	0.8	609	1	VG16_BPP22	Q01146 bacterioph
736	6	0.8	538	1	NPRB_BACSU	P39899 bacillus su	809	0.8	609	1	YL15_MYCLE	P46509 mycobacteri
737	6	0.8	540	1	G6PI_CORGL	Q9ns31 corynebacte	810	0.8	610	1	POUC_BRARE	P21367 brachydanio
738	6	0.8	540	1	GTR9_HUMAN	Q9nm0 homo sapien	811	0.8	611	1	ACES_FELCA	O62763 felis silve
739	6	0.8	541	1	60IM_HABIN	P44973 haemophilus	812	0.8	611	1	PHBC_RHIME	P50176 r poly-beta
740	6	0.8	542	1	DAGA_ALTHA	P30144 alteromonas	813	0.8	613	1	ACES_BOVIN	P22303 homo sapien
741	6	0.8	542	1	CHK2_HUMAN	Q96017 homo sapien	814	0.8	614	1	ACES_HUMAN	P21836 mus musculu
742	6	0.8	543	1	DALS_YEAST	P15365 saccharomyc	815	0.8	614	1	ACES_MOUSE	P37136 rattus norv
743	6	0.8	543	1	TIMD_MYCPN	P75436 mycoplasma	816	0.8	614	1	ACES_RAT	Q56235 thermus the
744	6	0.8	544	1	ALG6_YEAST	Q12001 saccharomyc	817	0.8	615	1	DNAX_THETH	P71773 mycobacteri
745	6	0.8	545	1	CH60_SODGL	Q9anr8 sodalis glo	818	0.8	615	1	MUTA_MYCTU	O04916 solanum tub
746	6	0.8	545	1	CATX_BACSU	P94377 bacillus su	819	0.8	616	1	RRPO_PLRV1	P17520 potato leaf
747	6	0.8	548	1	VG31_BPML5	Q05240 mycobacteri	820	0.8	616	1	Y396_HELPY	Q92kf3 helicobacte
748	6	0.8	550	1	COCH_HUMAN	Q43405 homo sapien	821	0.8	616	1	Y396_HELPY	O25157 helicobacte
749	6	0.8	550	1	MANB_MYCGE	P47299 mycoplasma	822	0.8	618	1	GIDA_CAUCR	Q9xbf8 caulobacter
750	6	0.8	551	1	ASLA_ECOLI	P25549 escherichia	823	0.8	618	1	ORC3_DROME	Q24168 drosophila
751	6	0.8	551	1	Y275_HABIN	P43975 haemophilus	824	0.8	618	1	GLMS_CYACA	O19908 c glucosami
752	6	0.8	552	1	AP1_SCHPO	Q01663 schizosacch	825	0.8	620	1	VP27_YEAST	P40343 saccharomyc
753	6	0.8	552	1	COCH_MOUSE	Q62507 mus musculu	826	0.8	622	1	PT07_YEAST	O28783 archaeoglob
754	6	0.8	552	1	VIE2_HSVB4	Q02479 bovine herp	827	0.8	623	1	PT07_YEAST	P51022 drosophila
755	6	0.8	555	1	HYFG_ECOLI	P77329 escherichia	828	0.8	623	1	PT07_YEAST	Q92jai helicobacte
756	6	0.8	557	1	ACH1_SCHGR	P23414 schistocerc	829	0.8	623	1	RECG_HELPY	O26051 helicobacte
757	6	0.8	557	1	AR11_HUMAN	Q9y4x5 homo sapien	830	0.8	624	1	STS_MOUSE	P50427 mus musculu
758	6	0.8	557	1	SCP2_MOUSE	Q9cun3 mus musculu	831	0.8	626	1	GLMS_NOS9	O68280 n glucosami
759	6	0.8	563	1	ARAB_MYCSM	Q91bq3 mycobacteri	832	0.8	628	1	HNFA_MOUSE	P22361 mus musculu
760	6	0.8	564	1	FRP1_SCHPO	Q4800 schizosacch	833	0.8	628	1	HNFA_MOUSE	P22361 mus musculu
761	6	0.8	564	1	OXC_ECOLI	P78093 escherichia	834	0.8	629	1	RA21_XENLA	P15257 rattus norv
762	6	0.8	564	1	SYT_MYCGE	P47615 mycoplasma	835	0.8	629	1	SYM_THEMA	O93310 xenopus lae
763	6	0.8	565	1	Y912_ARCFU	O29350 archaeoglob	836	0.8	629	1	SYM_THEMA	O33925 thermotoga

837	6	0.8	631	1	RA21_HUMAN	O60216	homo sapien	910	6	0.8	697	1	ATKB_RHILO	Q98gx6	rhizobium 1
838	6	0.8	632	1	T242_MTCPN	P75440	mycoplasma	911	6	0.8	698	1	GFAL_HUMAN	Q06210	homo sapien
839	6	0.8	633	1	HTPG_PSEAE	Q913c5	pseudomonas	912	6	0.8	699	1	PTGA_BACSU	P20166	bacillus su
840	6	0.8	634	1	HTPG_XANCP	Q8pjk3	xanthomonas	913	6	0.8	702	1	SPEI_ORYSA	Q9snn0	oryza sativ
841	6	0.8	635	1	HTPG_XANCP	Q8p855	xanthomonas	914	6	0.8	703	1	PALY_BROFI	Q42609	bromeadia
842	6	0.8	636	1	RA21_MOUSE	O61550	mus musculus	915	6	0.8	703	1	UL17_HSV11	P10201	herpes simp
843	6	0.8	637	1	DXS_COREF	O8fbi2	corynebacte	916	6	0.8	704	1	FBL1_CHICK	Q73775	gallus gall
844	6	0.8	639	1	ENV_FLVSA	P06752	feline leuk	917	6	0.8	707	1	ATKB_STRCO	Q9x8z9	streptomyce
845	6	0.8	641	1	DHAK_CITFR	P45512	citrobacter	918	6	0.8	709	1	ATKB_MYCTU	P96370	mycobacteri
846	6	0.8	641	1	SGRI_MOUSE	Q9c874	mus musculus	919	6	0.8	709	1	SYT_ARATH	O04630	arabidopsis
847	6	0.8	641	1	TETS_LISMO	O48791	listeria mo	920	6	0.8	710	1	PALI_RUBID	Q9m568	rubus idaeu
848	6	0.8	642	1	ENV_FLVGL	P08359	feline leuk	921	6	0.8	713	1	HS9B_HORSE	Q9mkx8	equus cabal
849	6	0.8	642	1	PRIM_MYCLE	O9ccg2	mycobacteri	922	6	0.8	713	1	PRML_BRARE	Q9W735	brachydanio
850	6	0.8	642	1	YB65_SCHPO	Q09746	schizosacch	923	6	0.8	716	1	HS83_DROAV	O02192	drosophila
851	6	0.8	643	1	SYT_EUCBP	P59554	buchnera ap	924	6	0.8	717	1	HS83_DROAV	P02828	drosophila
852	6	0.8	643	1	T2D4_SCHPO	O13282	schizosacch	925	6	0.8	718	1	ATKB_AGRT5	Q8u9d9	agrobacteri
853	6	0.8	644	1	BTD_DROME	Q24266	drosophila	926	6	0.8	718	1	KF2C_CRIGR	P70096	cricetulus
854	6	0.8	645	1	ENV_FSVSM	P21445	feline sarc	927	6	0.8	718	1	PNT2_DROME	P51023	drosophila
855	6	0.8	646	1	TETS_LACLA	O48712	lactococcus	928	6	0.8	719	1	HS9A_HORSE	Q95kx7	equus cabal
856	6	0.8	648	1	DPOI_BSP2	P06225	bacterioph	929	6	0.8	719	1	YM41_YEAST	Q03213	saccharomyc
857	6	0.8	648	1	MACB_ECOLI	P75831	escherichia	930	6	0.8	720	1	FPTA_PSEAE	P42512	pseudomonas
858	6	0.8	652	1	PICA_HUMAN	Q13492	homo sapien	931	6	0.8	720	1	HS78_SCHPO	O59838	schizosacch
859	6	0.8	654	1	YCO9_SCHPO	O13674	schizosacch	932	6	0.8	721	1	KF2C_MOUSE	Q92288	mus musculus
860	6	0.8	657	1	HUTH_MOUSE	P35492	mus musculus	933	6	0.8	722	1	HMN2_DROME	P22808	drosophila
861	6	0.8	657	1	HUTH_RAT	P21213	rattus norv	934	6	0.8	723	1	HS9B_HUMAN	P08238	homo sapien
862	6	0.8	658	1	SOHC_ZYMMO	P33990	zymomonas m	935	6	0.8	723	1	HS9B_MOUSE	P11499	mus musculus
863	6	0.8	659	1	RA24_YEAST	P32641	saccharomyc	936	6	0.8	723	1	HS9B_RAT	P14058	rattus norv
864	6	0.8	659	1	YK16_YEAST	P36130	saccharomyc	937	6	0.8	725	1	HS9B_BRARE	O57521	brachydanio
865	6	0.8	660	1	ALIA_STRPN	P35592	streptococc	938	6	0.8	725	1	HS9B_CHICK	O04619	gallus gall
866	6	0.8	660	1	DNLJ_BORBU	O51502	borrelia bu	939	6	0.8	725	1	PALY_TRISU	P45734	trifolium s
867	6	0.8	662	1	ACSA_HELPY	O25686	helicobacte	940	6	0.8	725	1	SYTC_CAEEL	P52709	caenorhabdi
868	6	0.8	663	1	OTRA_STRRM	O55002	streptomyce	941	6	0.8	728	1	HS9A_CHICK	P11501	gallus gall
869	6	0.8	665	1	DUSG_HUMAN	Q9by84	homo sapien	942	6	0.8	728	1	MASZ_BRUSE	Q8yir3	brucella me
870	6	0.8	669	1	TSEA_HUMAN	Q8wu66	homo sapien	943	6	0.8	728	1	MASZ_BRUSE	Q8fz50	brucella su
871	6	0.8	671	1	HMOC_DROME	P22810	drosophila	944	6	0.8	729	1	FLUE_ECOLI	P16869	escherichia
872	6	0.8	675	1	ATKB_DEIRA	Q9rxp0	deinococcus	945	6	0.8	730	1	GELS_HORSE	Q28372	equus cabal
873	6	0.8	676	1	AMV1_ECOLI	P25718	escherichia	946	6	0.8	731	1	HS9A_HUMAN	P07900	homo sapien
874	6	0.8	676	1	ATKB_ENTFA	O8ku73	enterococcu	947	6	0.8	732	1	EF2_PYRAB	Q9v128	pyrococcus
875	6	0.8	676	1	CHLD_SYNY3	P72772	synechocyst	948	6	0.8	732	1	EF2_PYRHO	O59521	pyrococcus
876	6	0.8	676	1	NTP2_VACCA	O57193	vaccinia vi	949	6	0.8	732	1	HS9A_CRIGR	P46633	cricetulus
877	6	0.8	676	1	NTP2_VACCC	P20502	vaccinia vi	950	6	0.8	732	1	HS9A_MOUSE	P07901	mus musculus
878	6	0.8	676	1	NTP2_VACCT	Q9jfc3	vaccinia vi	951	6	0.8	732	1	HS9A_PIG	O02705	sus scrofa
879	6	0.8	676	1	NTP2_VACCV	P12927	vaccinia vi	952	6	0.8	733	1	FLHA_HELPJ	Q92m40	helicobacte
880	6	0.8	676	1	NTP2_VARV	P33051	variola vir	953	6	0.8	733	1	FLHA_HELPJ	O06758	helicobacte
881	6	0.8	676	1	TIM_DROHY	O44331	drosophila	954	6	0.8	733	1	PTK_ACIOJ	O52788	acinetobact
882	6	0.8	677	1	UL17_VZVD	P09292	varicella-z	955	6	0.8	733	1	YACK_RHIME	Q9x447	rhizobium m
883	6	0.8	677	1	FLHA_BACSU	Q9bql6	homo sapien	956	6	0.8	735	1	CATA_BACST	P14412	bacillus st
884	6	0.8	678	1	NTP2_MXXVL	P35620	bacillus su	957	6	0.8	735	1	IF2C_GUITH	Q78489	guillardia
885	6	0.8	678	1	NTP2_SFVKA	Q9q8Q2	myxoma viru	958	6	0.8	736	1	VM1_REOVL	P12418	reovirus (t
886	6	0.8	678	1	VID3_AGRH	Q9q927	shope fibro	959	6	0.8	736	1	VM1_REOVL	Q00335	reovirus (t
887	6	0.8	679	1	PAN3_YEAST	P13463	agrobacteri	960	6	0.8	738	1	CHSY_USTMA	Q99127	ustilago ma
888	6	0.8	680	1	APB2_MOUSE	P36102	saccharomyc	961	6	0.8	740	1	GAG_SMRVH	P21411	squirrel mo
889	6	0.8	680	1	APB2_MOUSE	P98084	mus musculus	962	6	0.8	744	1	ATKB_RALSO	O8xul1	raistonia s
890	6	0.8	680	1	ATKB_RHIME	Q92xj0	rhizobium m	963	6	0.8	744	1	VG44_HALSO	P21562	haloferax s
891	6	0.8	681	1	ATB1_LISIN	Q927G0	listeria in	964	6	0.8	749	1	APB2_HUMAN	Q99767	homo sapien
892	6	0.8	681	1	ATKB_LISMO	O8y3z7	listeria mo	965	6	0.8	750	1	APB2_RAT	O35431	rattus norv
893	6	0.8	681	1	CRY1_ARATH	O43125	arabidopsis	966	6	0.8	751	1	EFGL_HUMAN	P54876	mycobacteri
894	6	0.8	682	1	ATKB_ECO57	Q8x9f5	escherichia	967	6	0.8	756	1	KP6F_CABEL	Q27483	caenorhabdi
895	6	0.8	682	1	ATKB_ECOLI	P03960	escherichia	968	6	0.8	758	1	YPS8_YEAST	Q92299	saccharomyc
896	6	0.8	682	1	ATKB_SALTY	O8zqw2	salmonella	969	6	0.8	763	1	PEPX_LACLA	Q9ce01	lactococcus
897	6	0.8	683	1	VTER_HGVSJ	Q01020	herpesvirus	970	6	0.8	766	1	PEPX_LACLC	P22346	lactococcus
898	6	0.8	684	1	FLID_HELPJ	Q9z191	helicobacte	971	6	0.8	766	1	PRTF_HSVBH	P28973	equine herp
899	6	0.8	684	1	ATKB_HELPY	P96786	helicobacte	972	6	0.8	769	1	COMP_BACSU	Q99027	bacillus su
900	6	0.8	686	1	ATKB_CAUCR	Q9axr7	caulobacter	973	6	0.8	769	1	MLH1_YEAST	P38920	saccharomyc
901	6	0.8	686	1	ATKB_MYXXA	Q9amv5	myxococcus	974	6	0.8	772	1	GELS_PIG	P20305	sus scrofa
902	6	0.8	688	1	ATKB_YERPE	Q8zd97	yersinia pe	975	6	0.8	775	1	PTNC_MOUSE	P35931	mus musculus
903	6	0.8	688	1	C1S_HUMAN	P09871	homo sapien	976	6	0.8	775	1	TNFS_MOUSE	Q60769	mus musculus
904	6	0.8	688	1	CACM_YEAST	P80235	saccharomyc	977	6	0.8	778	1	EXPT_STRPN	P35597	streptococc
905	6	0.8	688	1	ENV_MMTVB	P10259	mouse mamma	978	6	0.8	780	1	GELS_MOUSE	P13020	mus musculus
906	6	0.8	688	1	ENV_MMTVB	P03374	mouse mamma	979	6	0.8	780	1	MUS2_BORBU	O51125	borrelia bu
907	6	0.8	690	1	ATKB_PSEAE	P57698	pseudomonas	981	6	0.8	780	1	PTNC_HUMAN	Q05209	homo sapien
908	6	0.8	692	1	GLND_CORGL	Q9x706	corynebacte	982	6	0.8	781	1	GYRB_NEIGO	P22118	neisseria g
909	6	0.8	696	1	GFAL_MOUSE	P47856	mus musculus								

983 6 0.8 781 1 TL22_CHICK Qdgb6 gallus gall
 984 6 0.8 782 1 Q6396 homo sapien
 985 6 0.8 785 1 UBPI_HUMAN O94782 homo sapien
 986 6 0.8 790 1 KIF9_HUMAN O9haq2 homo sapien
 987 6 0.8 790 1 SYFB_CHIMU Q9pj18 chlamydia m
 988 6 0.8 790 1 TNP3_HUMAN P21580 homo sapien
 989 6 0.8 793 1 TL21_CHICK Q9d878 gallus gall
 990 6 0.8 797 1 D151_HAERIN P46024 haemophilus
 991 6 0.8 800 1 PT27_YEAST P32606 saccharomyc
 992 6 0.8 802 1 OPH1_HUMAN O60890 homo sapien
 993 6 0.8 802 1 OPH1_MOUSE Q99j31 mus musculu
 994 6 0.8 802 1 PF8C_SALTY P37868 salmonella
 995 6 0.8 802 1 PFN8_MOUSE P29352 mus musculu
 996 6 0.8 802 1 S15A_MOUSE Q9r313 mus musculu
 997 6 0.8 803 1 ATFCU_BAGSU O32220 bacillus su
 998 6 0.8 803 1 SYFB_WIGBR Q9d3b5 wiggleswort
 999 6 0.8 804 1 S15A_HUMAN Q8tag3 homo sapien
 1000 6 0.8 804 1 S15A_RAT O54923 rattus norv

ALIGNMENTS

RESULT 1

PE42_ARATH STANDARD; PRT; 330 AA.
 AC Q9GB81; Q41937; Q42579; Q43730;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Peroxidase 42 precursor (EC 1.11.1.7) (Atperox P42) (PRXR1)
 DE (ATP1a/ATP1b).
 GN PER42 OR P42 OR AT4G21960 OR TB05.170 OR FIN20.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia; TISSUE=Flower, Leaf, Root, Siliques, and Stem;
 RA Capelli N., Tognolli M., Flach J., Overney S., Penel C., Greppin H.,
 RA Simon P.;
 RA "Eleven cDNA clones from Arabidopsis thaliana encoding
 RT isoperoxidases.";
 RL (in) Plant Gene Register PGR96-066.
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=97238476; PubMed=9132061;
 RA Kjaergaard I.V.H., Jespersen H.M., Rasmussen S.K., Welinder K.G.;
 RT "Sequence and RT-PCR expression analysis of two peroxidases from
 RT Arabidopsis thaliana belonging to a novel evolutionary branch of plant
 RT peroxidases.";
 RL Plant Mol. Biol. 33:699-708(1997).
 RN [3]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoerg W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Billam L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,

De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 Pettett A., Rajandream M.A., Lyne M., Benes V., Rehnemann S.,
 Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
 Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
 Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 Schnabl S., Hiller R., Schmidt W., Lecharry A., Aubourg S.,
 Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 Heijman D., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 Stoneking T., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 Latreille P., Kalicki J., Graves T., Harmon G., Edwards J.,
 Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., L.,
 Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 Chen S., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 RN [4]
 RC SEQUENCE FROM N.A.
 RP Chao Q., Brooks S.Y., Chen H., Johnson-Hopson C., Khan S., Shinn P.,
 RA Ecker J.R.;
 RT "Full length cDNA sequence of Arabidopsis thaliana.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RC SEQUENCE OF 1-168 FROM N.A.
 RP TISSUE=Leaf;
 RA Stracke R., Palme K.;
 RT "Signal peptide selection derived cDNAs from Arabidopsis thaliana
 RT leaves and guard cells.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RC SEQUENCE OF 278-330 FROM N.A.
 RP STRAIN=cv. Columbia; TISSUE=Green siliques;
 RA Raynal M., Greillet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
 RT Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RC CHARACTERIZATION
 RP STRAIN=cv. Columbia;
 RX MEDLINE=98409321; PubMed=9738941;
 RA Oostergaard L., Pedersen A.G., Jespersen H.M., Brunak S.,
 RA Welinder K.G.;
 RT "Computational analyses and annotations of the Arabidopsis peroxidase
 RT gene family";
 RL FEBS Lett. 433:98-102(1998).
 RN [9]
 RC TISSUE SPECIFICITY.
 RP STRAIN=cv. Columbia;
 RA Zhu T., Budworth P., Han B., Brown D., Chang H.-S., Zou G., Wang X.;
 RT "Toward elucidating the global gene expression patterns of developing
 RT Arabidopsis: parallel analysis of 8300 genes by a high-density
 RT oligonucleotide probe array.";


```

FT METAL 247 247 CALCIUM 2 (BY SIMILARITY)
FT METAL 250 250 CALCIUM 2 (BY SIMILARITY)
FT METAL 255 255 CALCIUM 2 (BY SIMILARITY)
FT ACT_SITE 67 67 BY SIMILARITY
FT ACT_SITE 71 71 DISTAL HISTIDINE
FT ACT_SITE 167 167 SUBSTRATE BINDING (BY SIMILARITY)

Query Match 1.2%; Score 9; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
|||
D5 16 ALSATSEAV 24

RESULT 2
MY9B_RAT
ID MY9B_RAT STANDARD; PRT; 1980 AA.
AC Q63358;
LI 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin IxB (Unconventional myosin-9b).
DE MYO9B OR MYR5.
OS Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain stem, and Spinal cord;
RX MEDLINE=95188874; PubMed=7482973;
RA Reinhard J., Scheel A.A., Diekmann D., Hall A., Ruppert C.,
RA Baehler M.;
RT "A novel type of myosin implicated in signalling by rho family
RT GTPases.";
RL EMBO J. 14:697-704(1995).

CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC -!- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE
CC PERINUCLEAR REGION (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN TESTIS, LUNG, THYMUS, BRAIN,
CC LIVER, AND SPLEEN.
CC -!- SIMILARITY: Contains 4 IQ domains.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 Rho-associating domain.
CC -!- SIMILARITY: Contains 1 Ras-GAP domain.
CC -!- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG
CC binding domain.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
CC EMBL; X77609; CAA54700.1; --
CC PIR; S54307; S54307.
CC HSSP; P08799; 1WMD.
CC InterPro; IPR002219; DAG PE-bind.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR001609; myosin head.
CC InterPro; IPR000159; RA domain.
CC InterPro; IPR000198; Rho-GAP

```

```

DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF00612; IQ; 4.
DR Pfam; PF00063; myosin_head; 2.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00620; RhoGAP; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00324; RhoGAP; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50096; IQ; 3.
DR PROSITE; PS50200; RA; 1.
DR PROSITE; PS50238; RHO GAP; 1.
KW Myosin, Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;
KW Zinc.
FT DOMAIN 1 940 HEAD OR MOTOR DOMAIN.
FT DOMAIN 941 1045 NECK OR REGULATORY DOMAIN.
FT DOMAIN 1046 1380 TAIL.
FT DOMAIN 1841 1861 COILED COIL (POTENTIAL).
FT DOMAIN 1918 1948 COILED COIL (POTENTIAL).
FT DOMAIN 15 114 RAS-ASSOCIATING.
FT DOMAIN 845 856 ACTIN-BINDING.
FT DOMAIN 958 978 IQ 1.
FT DOMAIN 981 1001 IQ 2.
FT DOMAIN 1002 1024 IQ 3.
FT DOMAIN 1025 1054 IQ 4.
FT DOMAIN 1593 1641 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 1663 1848 RHO-GAP.
FT NP_BIND 239 246 ATP (POTENTIAL).
SQ SEQUENCE 1980 AA; 225035 MW; D79FEC4D0FAE0C05 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 1980;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSQ 213
DB 1690 TEGLYRKSQ 1698
|||||||

RESULT 3
ID MY9B MOUSE STANDARD; PRT; 2114 AA.
AC Q9QY06; Q9QY07; Q9QY08; Q9QY09;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Myosin IxB (Unconventional myosin-9b).
GN MY9B OR MYR5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANTS.
RC STRAIN=C57BL/6;
RX MEDLINE=20047919; PubMed=10580159;
RA Grewal P.K.; Jones A.-M.; Maconochie M.; Lemmers R.J.F.; Frants R.R.;
RA Hewitt J.E.;
RT "Cloning of the murine unconventional myosin gene Myo9b and
RT identification of alternative splicing.";
RL Gene 240:389-398(1999).
CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC

-!- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
LOCALIZATION IS CYTOSOLASMIC WITH THE HIGHEST LEVELS IN THE
PERINUCLEAR REGION (BY SIMILARITY).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Name=1;
IsoId=Q9QY06-1; Sequence=Displayed;
Name=2; Synonyms=Q;
IsoId=Q9QY06-2; Sequence=VSP_003363;
Name=3; Synonyms=C;
IsoId=Q9QY06-3; Sequence=VSP_003364; VSP_003365;
-!- TISSUE SPECIFICITY: EXPRESSED IN THE BRAIN, HEART, MUSCLE AND
INNER EAR.
-!- SIMILARITY: Contains 4 IQ domains.
-!- SIMILARITY: Contains 1 myosin-like globular head domain.
-!- SIMILARITY: Contains 1 Ras-associating domain.
-!- SIMILARITY: Contains 1 Rho-GAP domain.
-!- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG
binding domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; AF143687; AAF00122.1; -
EMBL; AF143685; AAF00120.1; -
EMBL; AF143686; AAF00121.1; ALT_FRAME.
EMBL; AF143683; AAF00118.1; -
HSP; P08799; 1MND.
MGD; MGI:106624; Myo9b.
InterPro; IPR002219; DAG_PE-bind.
InterPro; IPR000048; IQ_region.
InterPro; IPR001609; myosin_head.
InterPro; IPR00159; RA domain.
InterPro; IPR00198; RhoGAP.
Pfam; PF00130; DAG_PE-bind; 1.
Pfam; PF00612; IQ; 4.
Pfam; PF00063; myosin_head; 2.
Pfam; PF00788; RA; 1.
Pfam; PF00620; RhoGAP; 1.
PRINTS; PR00193; MYOSINHEAVY.
PRODOM; PD000355; myosin_head; 1.
SMART; SM00109; C1; 1.
SMART; SM00015; IQ; 4.
SMART; SM00242; MYSC; 1.
SMART; SM00314; RA; 1.
SMART; SM00324; RhoGAP; 1.
PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
PROSITE; PS50096; IQ; 3.
PROSITE; PS50200; RA; 1.
PROSITE; PS50238; RHO GAP; 1.
Myosin, Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;
Zinc; Alternative splicing; Polymorphism.
DOMAIN 1 939 HEAD OR MOTOR DOMAIN.
DOMAIN 940 1044 NECK OR REGULATORY DOMAIN.
DOMAIN 1045 2114 TAIL.
DOMAIN 1562 1592 COILED COIL (POTENTIAL).
DOMAIN 1839 1859 COILED COIL (POTENTIAL).
DOMAIN 1915 1945 COILED COIL (POTENTIAL).
DOMAIN 15 114 RAS-ASSOCIATING.
DOMAIN 844 855 ACTIN-BINDING.
DOMAIN 957 977 IQ 1.
DOMAIN 979 1000 IQ 2.
DOMAIN 1001 1023 IQ 3.
DOMAIN 1024 1053 IQ 4.
DOMAIN 1591 1639 PHORBOL-ESTER AND DAG BINDING.

```



```

FT DOMAIN 940 1044 NECK OR REGULATORY DOMAIN.
FT DOMAIN 1045 2158 TAIL.
FT DOMAIN 1046 2158 COILED COIL (POTENTIAL).
FT DOMAIN 1047 2158 COILED COIL (POTENTIAL).
FT DOMAIN 1880 1901 COILED COIL (POTENTIAL).
FT DOMAIN 1959 1989 COILED COIL (POTENTIAL).
FT DOMAIN 15 114 RAS-ASSOCIATING.
FT DOMAIN 844 855 ACTIN-BINDING.
FT DOMAIN 957 977 IQ 1.
FT DOMAIN 979 1000 IQ 2.
FT DOMAIN 1001 1023 IQ 3.
FT DOMAIN 1024 1053 IQ 4.
FT DOMAIN 1633 1681 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 1703 1888 RHO-GAP.
FT NP BIND 239 246 ATP (POTENTIAL).
FT VARSPLIC 2022 2023 PP -> QY (in isoform Short).
FT VARSPLIC 2022 2023 /FTid=VSP_003361.
FT VARSPLIC 2024 2158 Missing (in isoform Short).
FT VARSPLIC 2024 2158 /FTid=VSP_003362.
FT CONFLICT 1937 1939 QVP -> KT (IN REF. 4).
FT CONFLICT 1947 1947 L -> P (IN REF. 4).
FT CONFLICT 2040 2045 TVAAP -> PWPELH (IN REF. 3).
FT CONFLICT 2049 2049 P -> L (IN REF. 3).
FT CONFLICT 2067 2067 P -> S (IN REF. 3).
FT CONFLICT 2157 2158 NG -> MAESH (IN REF. 3).
SQ SEQUENCE 2158 AA; 243556 MW; 4978FID770F56D28 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 2158;
Best Local Similarity 100.0%; Pred.No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSQ 213
Db 1730 TEGLYRKSQ 1738
|||||

RESULT 5
YO96 AGRT5
ID YO96_AGR5 STANDARD; PRT; 196 AA.
AC P58792;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative NADH dehydrogenase/NAD(P)H nitroreductase Atu2496
DE (EC 1.1.1.17).
GN ATU2496 OR AGR_C_4536.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphammachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323 (2001).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,

Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001)
CC -1- COFACTOR: FMN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NITROREDUCTASE FAMILY. HADB SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE009196; AAL43483.1; -
DR EMBL; AE008162; AAK88228.1; -
DR PIR; AE2883; AE2883.
DR PIR; C97659; C97659.
DR HAMAP; MF 01204; -; 1.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
KW Hypothetical protein; Nitroreductase; NAD; NADP; Flavoprotein; FMN;
KW Complete proteome.
SQ SEQUENCE 196 AA; 21139 MW; 6283E17567796984 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 196;
Best Local Similarity 100.0%; Pred.No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 LSSGNLEK 187
Db 70 LSSGNLEK 77
|||||

RESULT 6
BLA3_BACCE
ID BLA3_BACCE STANDARD; PRT; 316 AA.
AC P06548;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Beta-lactamase III precursor (EC 3.5.2.6).
GN BLAZ.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=569/H;
RX MEDLINE=87109042; PubMed=3027036;
RA Hussain M., Pastor F.I.J., Lampen J.O.;
RT "Cloning and sequencing of the blaZ gene encoding beta-lactamase III,
RT a lipoprotein of Bacillus cereus 569/H."
RL J. Bacteriol. 169:579-586 (1987).
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M15195; AAA22274.1; -
DR PIR; A27755; PNEBLC.
DR HSP; P00808; 4BLM.
DR InterPro; IPR001466; Beta_lactamase.
```

```

DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BACTAMASE.
DR PROSITE: PS00013; PROKARYOTIC LIPOPROTEIN; 1.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolase; Antibiotic resistance; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 29
FT CHAIN 30 316 BETA-LACTAMASE III.
FT LIPID 30 30 N-ACYL DIGLYCERIDE.
FT ACT_SITE 95 95 BY SIMILARITY.
FT BINDING 257 259 SUBSTRATE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 35140 MW; 27A082AB9CCBC9DF CRC64;

Query Match 1.1%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KPVLVSIL 14
DB 278 KPVLVSIL 285

RESULT 7
FDH_PSESR
ID FDH_PSESR STANDARD; PRT; 400 AA.
AC P33160;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formate dehydrogenase (EC 1.2.1.2) (NAD-dependent formate
DE dehydrogenase) (FDH).
OS Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=33067;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92063805; PubMed=1954846;
RA Tishkov V.I., Galkin A.G., Egorov A.M.;
RT "NAD-dependent formate dehydrogenase of methylotrophic bacteria
RT Pseudomonas sp. 101: cloning, expression, and study of the genetic
RT structure.";
RL Dokl. Akad. Nauk SSSR 317:745-748 (1991).
RN [2]
RP SEQUENCE OF 1-393.
RX MEDLINE=90290536; PubMed=2357236;
RA Popov V.O., Shumilin I.A., Ustinikova T.B., Lamzin V.S., Egorov T.A.;
RT "NAD-dependent formate dehydrogenase from methylotrophic bacteria
RT Pseudomonas sp. 101. I. Amino acid sequence.";
RL Bioorg. Khim. 16:324-335 (1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=92283270; PubMed=1597184;
RA Lamzin V.S., Aleshin A.E., Strokopytov B.V., Yuhnevich M.G.,
RA Popov V.O., Harutyunyan E.H., Wilson K.S.;
RT "Crystal structure of NAD-dependent formate dehydrogenase.";
RL Eur. J. Biochem. 206:441-452 (1992).
RN [4]
RP MUTAGENESIS OF CYS-255.
RX MEDLINE=93249485; PubMed=8484798;
RA Tishkov V.I., Galkin A.G., Marchenko G.N., Egorova O.A., Sheluho D.V.,
RA Kulakova L.B., Dementieva L.A., Egorov A.M.;
RT "Catalytic properties and stability of a Pseudomonas sp.101 formate
RT dehydrogenase mutants containing Cys-255-Ser and Cys-255-Met
RT replacements.";
RL Biochem. Biophys. Res. Commun. 192:976-981 (1993).
CC -!- CATALYTIC ACTIVITY: Formate + NAD(+) = CO(2) + NADH.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY. STRONG, TO OTHER FDH.
DR PIR; J00334; J00334.
DR PDB; 2NAC; 26-JAN-95.
DR PDB; 2NAD; 26-JAN-95.
DR InterPro: IPR006139; 2-Hacid_DH.
DR InterPro: IPR006140; 2-Hacid_DH_C.
DR Pfam: PF00389; 2-Hacid_DH; 1.
DR Pfam: PF02826; 2-Hacid_DH_C; 1.
DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH_3; 1.
KW Oxidoreductase; NAD; 3D-structure.
FT INIT_MET 0 0
FT DOMAIN 1 146 CATALYTIC.
FT DOMAIN 147 333 COENZYME-BINDING.
FT DOMAIN 334 400 CATALYTIC.
FT NP_BIND 192 226 NAD (ADP PART).
FT ACT_SITE 5 5 POTENTIAL.
FT ACT_SITE 284 284 SUBSTRATE BINDING.
FT MUTAGEN 255 255 C->S,M: HIGH RESISTANCE TO INACTIVATION
BY HG(2+), INCREASED STABILITY AT 25 C.
AND DECREASED THERMOSTABILITY AT 45 C.
D -> S (IN REF. 1).
TV -> VT (IN REF. 1).
C -> V (IN REF. 1).
VH -> HV (IN REF. 1).
N -> D (IN REF. 1).

```


RA Paulin L., Perea J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,
RA Prydz H., Parnelle B., Rasmussen S.W., Renacha M., Revuelta J.L.,
RA Rieger M., Salom D., Saluz H.P., Saiz J.E., Saren A.-M., Schaefer M.,
RA Scharfe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,
RA Urrestazu L.A., Verhaselt P., Vissers S., Voet M., Volckaert G.,
RA Wagner G., Wambutt R., Wedler H., Woelfl S., Harris D.E.,
RA Bowman S., Brown D., Churher C.M., Connor R., Dedman K., Gentles S.,
RA Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,
RA Orell C., Oliver K., Rajandream M.A., Richards C., Shore L.,
RA Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen E.,
RA Araujo R., Aviles E., Berno A., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,
RA Lashkari D., Lew H., Lin D., Mosedale D., Nakahara K., Mamath A.,
RA Cefnegr P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,
RA Shogren T., Shroff N., Winant A., Telton M.A., Botstein D.,
RA Davis R.W., Johnston M., Andrews S., Brinkman R., Cooper J., Ding H.,
RA Du Z., Favell A., Fulton L., Gattung S., Greco T., Hallsworth K.,
RA Hawkins J., Hillier L., Jier M., Johnson D., Johnston L., Kirsten J.,
RA Kucaba T., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
RA Taich A., Trevisan E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R., Albermann K., Ham J., Heumann K., Kleine K.,
RA Mewes H.-W., Zollner A., Zaccaria P.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IV.";
RL Nature 387:75-78(1997).
CC -!- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -!- SIMILARITY: Contains 1 RHO-GAP domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U28373; AAB64815.1; -
CC HSSP; P27986; 1PBW
CC SGD; S0002787; RGA2.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR000198; RhoGAP.
CC Pfam; PF00412; LIM; 2.
CC ProDom; PD000094; RhoGAP; 1.
CC SMART; SM00132; LIM; 2.
CC SMART; SM00324; RhoGAP; 1.
CC PROSITE; PS00478; LIM DOMAIN 1; 1.
CC PROSITE; PS50023; LIM DOMAIN 2; 2.
CC PROSITE; PS50238; RHO-GAP; 1.
CC GTPase activation; Repeat; LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 11 68 LIM 1.
CC FT DOMAIN 69 129 LIM 2.
CC FT DOMAIN 788 1006 RHO-GAP.
CC SQ SEQUENCE 1009 AA; 113290 MW; FA13BA1DF8B35D98 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1009;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYKSG 213
Db 822 EGLYKSG 829

RESULT 10
NU3M DROME
ID NU3M DROME STANDARD; PRT; 117 AA.
AC P18930;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).

GN MT:ND3 OR ND3.
OS Drosophila melanogaster (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bretagne;
RX MEDLINE=88212147; PubMed=3130291;
RA Garesse R.;
RT "Drosophila melanogaster mitochondrial DNA: gene organization and
RT evolutionary considerations.";
RL Genetics 118:649-663(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R, and Zimbabwe 53;
RX MEDLINE=20363871; PubMed=10903372;
RA Ballard J.W.O.;
RT "Comparative genomics of mitochondrial DNA in members of the
RT Drosophila melanogaster subgroup.";
RL J. Mol. Evol. 51:48-63(2000).
RN [3]
RP SEQUENCE OF 1-29 FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=83220794; PubMed=6304652;
RA Clary D.O., Wahlthner J.A., Wolstenholme D.R.;
RT "Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
RT flanking sequences and comparisons to mammalian mitochondrial tRNA
RT genes.";
RL Nucleic Acids Res. 11:2411-2425(1983).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M37275; AAA69709.1; -
CC DR EMBL; U37541; AAC47817.1; -
CC DR EMBL; AF200828; AAF77232.1; -
CC DR EMBL; AF200829; AAF77244.1; -
CC DR EMBL; J01404; AAB59244.1; -
CC DR PIR; S01185; S01185.
CC DR FlyBase; FBgn0013661; mt:ND3.
CC DR InterPro; IPR000440; Oxidored_q4.
CC DR Pfam; PF00507; oxidored_q4; 1.
CC DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC KW SEQUENCE 117 AA; 13563 MW; 6DB7AFDA72924892 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 KALIDRE 467
Db 28 KALIDRE 34

RESULT 11
NU3M DROSU
ID NU3M DROSU STANDARD; PRT; 117 AA.
AC P51940;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN MT:ND3 OR ND3.


```

OS Drosophila subobscura (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP, TUE 3; PubMed=1454843;
RX MEDLINE=3087561; PubMed=1454843;
RA Volz-Lingenhoel A., Solignac M., Sperlich D.;
RT "Stable heteroplasmy for a large-scale deletion in the coding region
of Drosophila subobscura mitochondrial DNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11528-11532(1992).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65129; CAA46258.1; -.
DR FLYBase; FBgn0012957; Dsub\mt.ND3.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 117 AA; 13427 MW; 01FEDA1E94FB4012 CRC64;
-----
Query Match 0.9%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 461 KALIDRE 467
Db 28 KALIDRE 34
|||||

RESULT 12
ID NU3M DROYA STANDARD; PRT; 117 AA.
AC P07705;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN ND3.
OS Drosophila yakuba (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2317.6 Ivory Coast;
RX MEDLINE=86089137; PubMed=3001325;
RA Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial DNA molecular of Drosophila yakuba: nucleotide
sequence, gene organization, and genetic code.";
RL J. Mol. Evol. 22:252-271(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2317.6 Ivory Coast;
RX MEDLINE=84169539; PubMed=6324117;
RA Clary D.O., Wolstenholme D.R.;
RT "A cluster of six tRNA genes in Drosophila mitochondrial DNA that
includes a gene for an unusual tRNASerAGY.";
RL Nucleic Acids Res. 12:2367-2379(1984).
RN [4]
RP -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

```

```

CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03240; CAA26991.1; -.
DR EMBL; X00432; CAA25130.1; -.
DR PIR; G25797; G25797.
DR FLYBase; FBgn0013204; Dyak\mt.ND3.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 117 AA; 13457 MW; 3291F89E5DD1D19D CRC64;
-----
Query Match 0.9%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 461 KALIDRE 467
Db 28 KALIDRE 34
|||||

RESULT 13
ID PYRI_ECOLI STANDARD; PRT; 152 AA.
AC P00478;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartate carbamoyltransferase regulatory chain.
GN PYRI OR B4244 OR C5344 OR Z5855 OR ECS5221.
OS Escherichia coli,
OS Escherichia coli O6, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84119419; PubMed=6364131;
RA Schachman H.K., Pauza C.D., Navre M., Karels M.J., Wl L., Yang Y.R.;
RT "Location of amino acid alterations in mutants of aspartate
transcarbamoylase: structural aspects of interallelic
complementation.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:115-119(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [4]
RP SEQUENCE FROM N.A.

```

RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22(2001).
RN [6]
RP SEQUENCE OF 127-152 FROM N.A.
RX MEDLINE=86143826; PubMed=3912513;
RA Cunin R., Jacobs A., Charlier D.R.M., Crabeel M., Herve G.,
RA Glandsdorff N., Pierard A.;
RT "Structure-function relationship in allosteric aspartate
RT carbamoyltransferase from *Escherichia coli*. I. Primary structure of a
RT pyri gene encoding a modified regulatory subunit.";
RL J. Mol. Biol. 186:707-713(1985).
RN [7]
RP PRELIMINARY SEQUENCE.
RX MEDLINE=68284659; PubMed=4872216;
RA Weber K.;
RT "New structural model of *E. coli* aspartate transcarbamylase and the
RT amino-acid sequence of the regulatory polypeptide chain.";
RL Nature 218:1116-1119(1968).
RN [8]
RP SEQUENCE OF 1-10.
RC STRAIN-K12 / W3110;
RA Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA Hochstrasser D.F.;
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
RN [9]
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=79074799; PubMed=364472;
RA Monaco H.L., Crawford J.L., Lipscomb W.N.;
RT "Three-dimensional structures of aspartate carbamoyltransferase from
RT *Escherichia coli* and of its complex with cytidine triphosphate.";
RL Proc. Natl. Acad. Sci. U.S.A. 75:5276-5280(1978).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND SUBUNITS.
RX MEDLINE=84248054; PubMed=6377306;
RA Ke H., Honzatko R.B., Lipscomb W.N.;
RT "Structure of unligated aspartate carbamoyltransferase of *Escherichia*
RT coli at 2.6-A resolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040(1984).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=91104727; PubMed=2271528;
RA Stevens R.C., Gouaux J.E., Lipscomb W.N.;
RT "Structural consequences of effector binding to the T state of
RT aspartate carbamoyltransferase: crystal structures of the unligated
RT and ATP- and CTP-complexed enzymes at 2.6-A resolution.";
RL Biochemistry 29:7691-7701(1990).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=91104728; PubMed=2271529;
RA Gouaux J.E., Stevens R.C., Lipscomb W.N.;
RT "Crystal structures of aspartate carbamoyltransferase ligated with
RT phosphonoacetamide, malonate, and CTP or ATP at 2.8-A resolution and
RT neutral pH.";
RL Biochemistry 29:7702-7715(1990).
RN [14]
RP FUNCTION: INVOLVED IN ALLOSTERIC REGULATION OF ASPARTATE
RX CARBAMOYLTRANSFERASE.
CC -!- COFACTOR: Binds 1 zinc ion per subunit.
CC -!- SUBUNIT: HETERODECAMER (2C3:3R2) OF SIX CATALYTIC PYRB CHAINS
CC ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PYRI CHAINS
CC ORGANIZED AS THREE DIMERS (R2).
CC -!- SIMILARITY: BELONGS TO THE PYRI FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
RN [15]
RX EMBL; K01472; AAA24477.1; -;
DR EMBL; U14003; AAA97141.1; -;
DR EMBL; AF000495; AAC77201.1; -;
DR EMBL; AF016771; AAN83765.1; -;
DR EMBL; AF005657; BAG59442.1; -;
DR EMBL; AF002568; BAB38644.1; -;
DR EMBL; M28578; AAA24487.1; -;
DR PIR; A93985; DTECR.
DR PIR; E91281; E91281.
DR PIR; F86122; F86122.
DR PDB; 1ACM; 15-JUL-92.
DR PDB; 2ATC; 15-JUL-92.
DR PDB; 8ATC; 15-OCT-90.
DR PDB; 1AT1; 15-OCT-90.
DR PDB; 2AT1; 15-OCT-90.
DR PDB; 3AT1; 15-OCT-90.
DR PDB; 4AT1; 15-OCT-90.
DR PDB; 5AT1; 31-JUL-94.
DR PDB; 6AT1; 15-OCT-90.
DR PDB; 7AT1; 15-OCT-90.
DR PDB; 8AT1; 31-JUL-94.
DR PDB; 1RAA; 31-JAN-94.
DR PDB; 1RAB; 31-JAN-94.
DR PDB; 1RAC; 31-JAN-94.
DR PDB; 1RAD; 31-JAN-94.
DR PDB; 1RAE; 31-JAN-94.
DR PDB; 1RAF; 31-JAN-94.
DR PDB; 1RAG; 31-JAN-94.
DR PDB; 1RAH; 31-JAN-94.
DR PDB; 1RAI; 31-JAN-94.
DR PDB; 1RBE; 14-OCT-98.
DR PDB; 1D09; 23-MAY-00.
DR PDB; 1EZZ; 22-NOV-00.
DR PDB; 1F1B; 22-NOV-00.
DR PDB; 1I5O; 02-MAY-01.
DR SWISS-2DPAGE; P00478; COLI.
DR ECODDBASE; I017.5; 6TH EDITION.
DR Ecogene; EGI0811; pyri.
DR HAMAP; MF_00002; -; 1.
DR InterPro; IPR002801; Pyri.
DR Pfam; PF01948; Pyri; 1.
DR Pfam; PF02748; Pyri_C; 1.
DR ProDom; PD006194; Pyri; 1.
DR TIGRFAMs; TIGR00340; ATCase reg; 1.
KW Pyrimidine biosynthesis; Zinc; 3D-structure; Complete proteome.
FT INIT MET 0
FT METAL 108 108 ZINC.
FT METAL 113 113 ZINC.


```

RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC glyceraldehyde 3-phosphate + acetaldehyde.
CC -!- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DEOC/FRAB FAMILY OF ALDOLASES. DEOC
CC SUBFAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AE005086; AAG20062.1; -.
DR PIR; B84337; B84337.
DR HAMAP; MF 00114; -.
DR InterPro; IPR002915; Deoc.
DR InterPro; IPR003009; FMN_enzyme.
DR Pfam; PF01791; Deoc; 1.
KW Lyase; Schiff base; Complete proteome.
FT BINDING 151 151
FT SCHIFF-BASE WITH ACETALDEHYDE (BY
FT SIMILARITY).
FT ACT SITE 176 176
FT BY SIMILARITY.
SQ SEQUENCE 212 AA; 21472 MW; F9DCF56817DE87DA CRC64;

Query Match 0.9%; Score 7; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 PIGRLKG 395
DB 94 PIGRLKG 100
|||||

RESULT 19
Y116 METJA
ID Y116 METJA STANDARD; PRT; 225 AA.
AC Q57580;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0116.
GN MJ0116.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

```

```

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U67469; AAB98097.1; -.
DR PIR; D64314; D64314.
DR TIGR; MJ0116; -.
DR InterPro; IPR003785; Creatininase.
DR Pfam; PF02633; Creatininase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25305 MW; 929EF2CE84DE35C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 LRLSSGN 184
DB 3 LRLSSGN 9
|||||

RESULT 20
FA3A HUMAN
ID FA3A HUMAN STANDARD; PRT; 230 AA.
AC P98173; Q96H51;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein FAM3A precursor (2-19 protein).
GN FAM3A OR 2-19 OR 2.19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=94068527; PubMed=8248200;
RA Bione S., Tamani F., Maestrini E., Tribioli C., Poustka A.,
RA Torri G., Rivella S., Toniolo D.;
RT "Transcriptional organization of a 450-kb region of the human X
RT chromosome in Xq28.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10977-10981 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Zollo M., Mazzarella R., Bione S., Toniolo D., Schlessinger D.,
RA D'Urso M., Chen E.Y.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96311563; PubMed=8733135;
RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: thirteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RL Hum. Mol. Genet. 5:659-668 (1996).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANT TRP-213.
RX TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=22150867; PubMed=12160727;
RA Zhu Y., Xu G., Patel A., McLaughlin M.M., Silverman C., Knecht K.A.,
RA Sweitzer S., Li X., McDonnell P., Mirabile R., Zimmerman D., Boyce R.,
RA Tierney L.A., Hu E., Livi G.P., Wolf B.A., Abdel-Meguid S.S.,
RA Rose G.D., Aurora R., Hensley P., Briggs M., Young P.R.;
RT "Cloning, expression, and initial characterization of a novel
RT cytokine-like gene family."
RL Genomics 80:144-150(2002).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- TISSUE SPECIFICITY: IN SIMILAR AMOUNTS IN TESTIS, PANCREAS,
CC ADRENAL, PLACENTA, BRAIN, FETAL BRAIN, LIVER, KIDNEY, SKELETAL
CC MUSCLE, HEART.
CC -!- SIMILARITY: BELONGS TO THE FAM3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X55448; CAA39090.1; -;
DR EMBL; X87193; CAA60645.1; -;
DR EMBL; L44140; AAA92652.1; -;
DR EMBL; BC008912; AAH08912.1; -;
DR EIR; 137095; I37095.
DR Genew; HGNC:13749; FAM3A.
KW Signal; Polymorphism.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 230 PROTEIN FAM3A.
FT DISULFID 59 222 POTENTIAL.
FT DISULFID 65 87 POTENTIAL.
FT VARIANT 213 213 C->W (IN dbSNP:1050788).
SQ SEQUENCE 230 AA; 25069 MW; FE3934D91F98CAAD CRC64;
FTID=VAR_011923.
Query Match 0.9%; Score 7; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 170 NGVSGEL 176
Db 111 NGVSGEL 117
RESULT 21
FA3A_MOUSE
ID FA3A_MOUSE STANDARD; PRT; 230 AA.
AC Q9D870;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein FAM3A precursor.

GN FAM3A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO THE FAM3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AK007715; BAB25208.1; -;
DR MGD; MGI:1913544; 1810037C20Rik.
KW Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 230 PROTEIN FAM3A.
FT DISULFID 59 222 POTENTIAL.
FT DISULFID 65 87 POTENTIAL.
SQ SEQUENCE 230 AA; 25393 MW; 6239418ED8258087 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 170 NGVSGEL 176
Db 111 NGVSGEL 117
RESULT 22
KC3A_MOUSE
ID KC3A_MOUSE STANDARD; PRT; 231 AA.
AC Q9EWW6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytidylylate kinase (EC 2.7.4.14) (CK) (Cytidine monophosphate kinase)
DE (CMP kinase).
GN CMK OR SCOL760 OR 2SCI34.13C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]

```
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -I- CATALYTIC ACTIVITY: ATP + (d)CMP = ADP + (d)CDP.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE CYTIDYLATE KINASE FAMILY. SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL939110; CAC12933.1; -.
DR HSSP; P23863; 1CKE.
DR HAMAP; MF 00238; 1.
DR InterPro; IPR003136; Cytidylate_kin.
DR TrEMBL; P02224; Cytidylate_kin; 1.
DR TIGRFAMs; TIGR00017; cmk; 1.
DR Transferase; Kinase; ATP-binding; Complete proteome.
FT NP BIND 18 26 ATP (BY SIMILARITY).
FT SEQUENCE 231 AA; 23600 MW; 76AC463B6F5BFD7E CRC64;
Query Match 0.9%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 365 KAGDAVE 371
Db 199 KAGDAVE 205
RESULT 23
ID ULA3 HCMVA STANDARD; PRT; 249 AA.
AC P16734; 1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Protein UL103.
GN UL103.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Reddell E., Satchwell S.C., Tomlinson P., Weston K.M., Barrall B.G.;
RA "Analysis of the protein-coding content of the sequence of human
RA cytomegalovirus strain AD169.";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -I- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,
CC EHV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HVS-1 42.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; CAA35339.1; -.
DR PIR; S09868; S09868.
DR InterPro; IPR002600; Herpes UL7.
DR Pfam; PF01677; Herpes UL7; 1.
DR SEQUENCE 249 AA; 28636 MW; A8D9F8F9F02FE9D CRC64;
Query Match 0.9%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 723 VKADNLL 729
Db 95 VKADNLL 101
RESULT 24
ID YE94 XYLFA STANDARD; PRT; 281 AA.
AC Q9PD85;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein XF1494.
GN XF1494.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferri J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
RA Vallada H., Van Sluys M.A., Vertovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000)
CC -I- SIMILARITY: BELONGS TO THE UPF0162 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

CC -----
DR EMBL; AF003979; AAF84303.1; ALT_INIT.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 281 AA; 32189 MW; 47D8879D22181684 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 281;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 RYLQLYP 292

DB 252 RYLQLYP 258

|||||

RESULT 25

HEM3_RHILO

ID HEM3_RHILO STANDARD; PRT; 308 AA.

AC Q98E17;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Porphobilinogen deaminase (EC 4.3.1.8) (PBG) (Hydroxymethylbilane

DE synthase) (HMB) (Pre-uroporphyrinogen synthase).

GN HEMC OR MLL4223.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI_TaxID=381;

[1]

SEQUENCE FROM N.A.

RC STRAIN=MAFF303099;

RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,

RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti.";

RL DNA Res. 7:331-338(2000).

CC -I- FUNCTION: Tetrapolymerization of the monopyrrole PBG into the

CC hydroxymethylbilane preuroporphyrinogen in several discrete steps.

CC -I- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =

CC hydroxymethylbilane + 4 NH(3).

CC -I- COFACTOR: Covalently binds a dipyrromethane cofactor to which the

CC porphobilinogen subunits are added (By similarity).

CC -I- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.

CC -I- SUBUNIT: Monomer (By similarity).

CC -I- SIMILARITY: BELONGS TO THE HMB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC -----
DR HEMAP; AF003003; BAB50931.1; -.

DR InterPro; IPR00860; Porphobil deam.

DR Pfam; PF01379; Porphobil deam; 1.

DR Pfam; PF03900; Porphobil deamC; 1.

DR PRINTS; PR00151; PORPHBDMNASE.

DR ProDom; PD002745; Porphobil deam; 1.

DR TIGRFAMS; TIGR00212; hemC; 1.

DR PROSITE; PS00533; PORPHOBILINOGEN DEAM; 1.

KW Porphyrin biosynthesis; Lyase; Complete proteome.

FT BINDING 243 243 PYRROMETHANE COFACTOR (BY SIMILARITY).

SQ SEQUENCE 308 AA; 32579 MW; 8164DE3AAFBE442 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 308;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LAQAHEH 27

DB 16 LAQAHEH 22

|||||

RESULT 26

TPIC_ARATH

ID TPIC_ARATH STANDARD; PRT; 315 AA.

AC Q9SKF6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Triosephosphate isomerase, chloroplast precursor (EC 5.3.1.1) (TIM).

GN TIM OR AT2G21170 OR F26H11.7.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

[1]

SEQUENCE FROM N.A.

RA Shih M.-C.;

RT "Structure and regulation of nuclear genes encoding chloroplast and

RT cytosolic triosephosphate isomerase in Arabidopsis thaliana.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=2083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shee T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Rinning C.M., Koo H.L.,

RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,

RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,

RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,

RA Nierman W.C., White O., Bisen J.A., Salzberg S.L., Fraser C.M.,

RA Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

RT thaliana.";

RL Nature 402:761-768(1999).

CC -I- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone

CC phosphate.

CC -I- PATHWAY: Calvin cycle.

CC -I- SUBUNIT: Homodimer (By similarity).

CC -I- SUBCELLULAR LOCATION: Chloroplast (By similarity).

CC -I- MISCELLANEOUS: IN PLANTS, THERE ARE TWO TYPES OF TPIS, CYTOSOLIC

CC AND PLASTID.

CC -I- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC -----
DR EMBL; AF006264; AAD29799.1; -.

DR PIR; AF247559; AAF70259.1; -.

DR PIR; AB4598; AB4598.

DR HSP; P00940; ITPH.

DR InterPro; IPR000652; Triophos_ismrse.

DR Pfam; PF00121; TIM; 1.

DR ProDom; PD001005; Triophos_ismrse; 1.

DR TIGRFAMS; TIGR00419; tim; 1.

DR PROSITE; PS00171; TIM; 1.

KW Isomerase; Chloroplast; transit peptide; Calvin cycle.

```
FT TRANSIT 1 60 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 61 315 TRIOSEPHOSPHATE ISOMERASE.
FT ACT SITE 155 155 BY SIMILARITY.
FT ACT_SITE 225 225 BY SIMILARITY.
SQ SEQUENCE 315 AA; 33345 MW; BD580F9B37FA0269 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 671 SLTDRID 677
DB 114 SLTDRID 120

RESULT 27
FLGJ VIBCH STANDARD; PRT; 334 AA.
AC Q9KQ15; O30855;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Peptidoglycan hydrolase flgJ (EC 3.2.1.-) (Muramidase flgJ).
GN FLGJ OR VC2192.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor Inaba V86;
RX MEDLINE=98415119; PubMed=9742694;
RA Das M., Chopra A.K., Wood T., Peterson J.W.;
RT "Cloning, sequencing and expression of the flagellin core protein and
RT other genes encoding structural proteins of the Vibrio cholerae
RT flagellum".
RL FEMS Microbiol. Lett. 165:239-246 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.B., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae".
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: FLAGELLUM-SPECIFIC MURAMIDASE WHICH HYDROLYZES THE
CC PEPTIDOGLYCAN LAYER TO ASSEMBLE THE ROD STRUCTURE IN THE
CC PERIPLASMIC SPACE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- MISCELLANEOUS: PROBABLY EXPORTED VIA THE FLAGELLUM-SPECIFIC EXPORT
CC PATHWAY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC GLYCOSYL HYDROLASES.
CC -!- SIMILARITY: BELONGS TO THE FLGJ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF019213; AAC33157.1; --
CC EMBL; AE004291; AAF95337.1; --
CC FIC; G82105; G82105.
CC TIGR; VC2192; --
```

```
DR InterPro; IPR002901; Amidase 4.
DR InterPro; IPR00423; Flag FlgJ.
DR Pfam; PF01832; Amidase 4; 1.
DR PRINTS; PR01002; FLGFLGJ.
DR SMART; SM00047; LY22; 1.
KW Flagella; Hydrolase; Glycosidase; Cell wall; Periplasmic;
KW Complete proteome.
FT DOMAIN 160 334 CATALYTIC.
FT ACT_SITE 231 231 POTENTIAL.
FT ACT_SITE 256 256 POTENTIAL.
FT CONFLICT 90 90 S -> C (IN REF. 1).
FT CONFLICT 312 334 YKSPAGIFRRSFSIFSAYQS -> L (IN REF. 2).
SQ SEQUENCE 334 AA; 37271 MW; DEA90DB0745450A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LRQKAVN 72
DB 20 LRQKAVN 26

RESULT 28
XYNB STRLI STANDARD; PRT; 335 AA.
AC P26515;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endo-1,4-beta-xylanase B precursor (EC 3.2.1.8) (Xylanase B)
DE (1,4-beta-D-xylan xylanohydrolase B).
GN XLNB.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 42-72.
RC STRAIN=66 / 1326;
RX MEDLINE=92077439; PubMed=1743521;
RA Shareck F., Roy C., Yaguchi M., Morosoli R., Kluepfel D.;
RT "Sequences of three genes specifying xylanases in Streptomyces
RT lividans".
RL Gene 107:75-82 (1991).
RN [2]
RP REVISIONS TO 29-32 AND 252-307.
RC STRAIN=66 / 1326; PubMed=7533741;
RX MEDLINE=95189090; PubMed=7533741;
RA Shareck F., Biely P., Morosoli R., Kluepfel D.;
RT "Analysis of DNA flanking the xlnB locus of Streptomyces lividans
RT reveals genes encoding acetyl xylan esterase and the RNA component of
RT ribonuclease P".
RN Gene 153:105-109 (1995).
RN [3]
RP REVISION TO 225.
RA Shareck F.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Contributes to hydrolyze hemicellulose, the major
CC component of plant cell-walls. XLNA and XLNB seem to act
CC sequentially on the substrate to yield xylobiose and xylose
CC as carbon sources.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC linkages in xylans.
CC -!- PATHWAY: Xylan degradation.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY G (FAMILY 11 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M64552; AAC06114.2; -.
DR HSP; P09850; IXNB.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR001137; Glyco_hydro_11.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR PRINTS; PR00911; GLHYDRLASE11.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
KW Xylan degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 41
FT CHAIN 42 335
FT DOMAIN 42 230 ENDO-1,4-BETA-XYLANASE B.
FT DOMAIN 231 249 CATALYTIC.
FT DOMAIN 250 335 LINKER ("HINGE") (GLY-RICH BOX).
FT DOMAIN 250 335 XYLAN-BINDING (POTENTIAL).
FT ACT_SITE 128 128 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 218 218 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 335 AA; 35575 MW; 513B1458BF0FF0CF CRC64;

Query Match 0.9%; Score 7; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 RGPVTL 146
Db 12 RGPVTL 18

RESULT 29
TRM1_ARCFU STANDARD; PRT; 349 AA.
ID TRM1_ARCFU
AC Q29443;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N(2),N(2)-dimethylguanosine tRNA methyltransferase (EC 2.1.1.32)
DE (tRNA [guanine-26,N(2)-N(2)] methyltransferase) (tRNA 2,2-
DE dimethylguanosine-26 methyltransferase)
DE (tRNA (m(2,2)G26)dimethyltransferase).
GN TRM1 OR AF0815.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervagge A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Dimethylates a single guanine residue at position 26 of
CC a number of tRNAs using S-adenosyl-L-methionine as donor of the
CC methyl groups (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing N(2)-methylguanine.
CC -1- SIMILARITY: BELONGS TO THE TRM1 FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE001048; AAB90427.1; -.
DR PIR; G69351; G69351.
DR TIGR; AF0815; -.
DR HAMAP; MF 00290; -. 1.
DR InterPro; IPR002905; TRM.
DR Pfam; PF02005; TRM; 1.
DR TIGRFAMs; TIGR00308; TRM1; 1.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
SQ SEQUENCE 349 AA; 38808 MW; 963859AFC18987F9 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 ENGVSGE 175
Db 81 ENGVSGE 87

RESULT 30
MLTA_VIBCH STANDARD; PRT; 368 AA.
ID MLTA_VIBCH
AC Q2KPQ4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-)
DE (Murein hydrolase A).
GN MLTA OR VC2312.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: MUREIN-DEGRADING ENZYME. MAY PLAY A ROLE IN RECYCLING OF
CC MUROPETIDES DURING CELL ELONGATION AND/OR CELL DIVISION. DEGRADES
CC MUREIN GLYCAN STRANDS AND INSOLUBLE, HIGH-MOLECULAR WEIGHT MUREIN
CC SACculi (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Cleavage of the beta-1,4-glycosidic bond
CC between N-acetylmuramic acid and N-acetylglucosamine residues,
CC thereby conserving the energy in a newly synthesized
CC 1,6-anhydrobond in the muramic acid residue.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (By similarity).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```
CC ENBL; AE004301; AAF95456.1; -.
DR DR PIR; A82093; A82093.
DR TIGR; VC2312; -.
DR InterPro; IPR005300; Mita.
DR Pfam; PF03562; Mita; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Cell wall; Hydrolase; Glycosidase; Signal; Lipoprotein;
FT Outer membrane; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 368
FT LIPID 17 17
FT SEQUENCE 368 AA; 40458 MW; C1341A9D21A41DBB CRC64;
Query Match 0.9%; Score 7; DB 1; Length 368;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 393 IELRHQP 389
Db 121 IELRHQP 127
RESULT 31
HEM3_PEA STANDARD; PRT; 369 AA.
AC Q43082;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Porphobilinogen deaminase, chloroplast precursor (EC 4.3.1.8) (PBG)
DE (Hydroxymethylbilane synthase) (HMS) (Pre-uroporphyrinogen synthase).
GN HEMC.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=cv. Little Marvel; TISSUE=Leaf;
RX MEDLINE=94269188; PubMed=7516080;
RA Witty M., Wallace-Cook A.D.M., Albrecht H., Spano A.J., Michel H.,
RA Shabanowitz J., Hunt D.F., Tinko M.P., Smith A.G.;
RT "Structure and expression of chloroplast-localized porphobilinogen
RT deaminase from pea (Pisum sativum L.) isolated by redundant
RT polymerase chain reaction."
RL Plant Physiol. 103:139-147(1993).
CC -!- FUNCTION: TETRAPOLYMERIZATION OF THE MONOPYRROLE PBG INTO THE
CC -!- HYDROXYMETHYLBILANE PREUROPORPHYRINOGEN IN SEVERAL DISCRETE STEPS.
CC -!- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =
CC hydroxymethylbilane + 4 NH(3).
CC -!- COFACTOR: COVALENTLY BINDS A DIPYRROMETHANE COFACTOR TO WHICH
CC THE PORPHOBILINOGEN SUBUNITS ARE ADDED.
CC -!- PATHWAY: porphyrin biosynthesis by the C5 pathway; fourth step.
CC Involved in chlorophyll biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE HMS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X73418; CAA51820.1; -.
DR PIR; S35873; JQ2278.
DR HSPF; P06983; 2YFN.
DR InterPro; IPR000860; Porphobil deam.
DR Pfam; PF01379; Porphobil deam; 1.
DR Pfam; PF03900; Porphobil deamC; 1.
```

```
DR PRINTS; PR00151; PORPHEDMNASE.
DR ProDom; PD002745; Porphobil deam; 1.
DR TIGRFAMS; TIGR00212; hemC; 1.
DR PROSITE; PS00533; PORPHOBILINOGEN DEAM; 1.
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Lyase; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 46
FT CHAIN 47 369
FT BINDING 303 303
FT SEQUENCE 369 AA; 39930 MW; F7660D8390894431 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 369;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 21 LAQAHEH 27
Db 73 LAQAHEH 79
RESULT 32
ALR_MYCLE STANDARD; PRT; 388 AA.
AC P38056;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN ALR OR ML0375 OR B229 C3_243.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Provides the D-alanine required for cell wall
CC biosynthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; first
CC step.
CC
CC -!- SIMILARITY: Belongs to the alanine racemase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U00020; AAA17309.1; -.
DR EMBL; AL583918; CAC29883.1; -.
DR PIR; S72995; S72995.
DR HSPF; P10724; 1BD0.
DR Leproma; ML0375; -.
```

```

DR HAMAP; MF_01201; -. 1.
DR InterPro; IPR000821; Ala_racemase.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR TIGRFAMs; TIGR00492; alr; 1.
DR PROSITE; PS00395; ALANINE_RACEMASE; 1.
KW Isomerase; Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis;
KW Complete proteome.
FT ACT_SITE 44 44 CATALYTIC BASE SPECIFIC TO D-ALANINE (BY
FT ACT_SITE 44 44 SIMILARITY).
FT ACT_SITE 273 273 CATALYTIC BASE SPECIFIC TO L-ALANINE (BY
FT ACT_SITE 273 273 SIMILARITY).
FT BINDING 44 44 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 388 AA; 41083 MW; 84075B683D00D196 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LLAQAE 26
Db 196 LLAQAE 202

RESULT 33
PEPF RABIT
ID PEPF RABIT STANDARD; PRT; 388 AA.
AC P27823;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pepsin F precursor (EC 3.4.23.1).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: hydrophobic, preferably
CC aromatic, residues in P1 and P1' positions. Cleaves 1-Phe-|-Val-2,
CC 4-Gln-|-His-5, 13-Glu-|-Ala-14, 14-Ala-|-Leu-15, 15-Leu-|-Tyr-16,
CC 16-Tyr-|-Leu-17, 23-Gly-|-Phe-24, 24-Phe-|-Phe-25 and 25-Phe-|-
CC Tyr-26 bonds in the B chain of insulin.
CC -1- DEVELOPMENTAL STAGE: EARLY POSTNATAL.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC -----
DR M59238; AAA31440.1; -.
DR HSP; P00794; 4CWS.
DR MEROPS; A01.051; -.
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR001461; AsproteaseA1.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PREPSIN.

```

```

DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 58 PEPSIN F.
FT CHAIN 59 388 BY SIMILARITY.
FT ACT_SITE 92 92 BY SIMILARITY.
FT ACT_SITE 275 275 BY SIMILARITY.
FT DISULFID 105 110 BY SIMILARITY.
FT DISULFID 266 270 BY SIMILARITY.
FT DISULFID 309 343 BY SIMILARITY.
SQ SEQUENCE 388 AA; 42786 MW; 24792BE393594B3A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
Db 152 AGLVDVA 158

RESULT 34
A23D DROME
ID A23D DROME STANDARD; PRT; 398 AA.
AC Q24093; Q9VOM6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CG3488 protein.
GN CG3488 OR ANON-23DA OR ANON-23D.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Chertoff E.H., Gelbart W.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.Y.,
RA Brandon R.C., Rogers Y.-H.C., Blaese R.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Heston D., Hovet D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

Fahy J., Helton E., Kettenan M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
Schnersch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."; *Proc. Natl. Acad. Sci. U.S.A.* 99:16899-16903 (2002).
[3]
SEQUENCE OF 6-429 FROM N.A.
MEDLINE=87134779; PubMed=2434381;
Leube R.E., Bosch F.X., Romano V., Zimbelmann R., Hofler H.,
Franke W.W.;
"Cyokeratin expression in simple epithelia. III. Detection of mRNAs
encoding human cytokeratins nos. 8 and 18 in normal and tumor cells
by hybridization with cDNA sequences in vitro and in situ."; *Differentiation* 33:69-85 (1986).
[4]
SEQUENCE OF 198-429 FROM N.A.
MEDLINE=86193258; PubMed=2422083;
Romano V., Harzfeld M., Magin T.M., Zimbelmann R., Franke W.W.,
Maier G., Ponattingi H.;
"Cyokeratin expression in simple epithelia. I. Identification of
mRNA coding for human cytokeratin no. 18 by a cDNA clone."; *Differentiation* 30:244-253 (1986).
[5]
SEQUENCE OF 1-166 FROM N.A.
MEDLINE=88246424; PubMed=2454392;
Kulesh D.A., Oshima R.G.;
"Cloning of the human keratin 18 gene and its expression in
nonepithelial mouse cells."; *Mol. Cell. Biol.* 8:1540-1550 (1988).
[6]
PARTIAL SEQUENCE.
TISSUE=Colon carcinoma;
MEDLINE=977295306; PubMed=9150948;
Ji H., Reid G.B., Moritz R.L., Edges J.S., Burgess A.W., Simpson R.J.;
"A two-dimensional gel database of human colon carcinoma proteins."; *Electrophoresis* 18:605-613 (1997).
[7]
CARBOHYDRATE-LINKAGE SITES.
MEDLINE=95263524; PubMed=7538124;
Xu N.-O., Omary M.B.;
"Identification and mutational analysis of the glycosylation sites of
human keratin 18."; *J. Biol. Chem.* 270:11820-11827 (1995).
[8]
VARIANT CC LEU-127.
MEDLINE=97148766; PubMed=9011570;
Xu N.-O., Wright T.L., Terrault N.A., Gish R., Omary M.B.;
"Mutation of human keratin 18 in association with cryptogenic
cirrhosis."; *J. Clin. Invest.* 99:19-23 (1997).
-!- SUBUNIT: HETEROOTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
-!- KERATIN 18 ASSOCIATES WITH KERATIN 8.
-!- DISEASE: DEFECTS IN KRT18 ARE TO BE A PREDISPOSITION OF
CYTOGENIC CIRRHOSIS (CC).
-!- MISCELLANEOUS: There are two types of cytoskeletal and
microfibrillar keratin, I (acidic) and II (neutral to basic) (40-
55 and 56-70 kDa, respectively).
-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AF179904; AAA59461.1; -
EMBL; BC000180; AAA00180.1; -
DR

```

DR EMBL; BC000698; AAH00698.1; -.
DR EMBL; BC004253; AAH04253.1; -.
DR EMBL; BC008636; AAH08636.1; -.
DR EMBL; BC020982; AAH20982.1; -.
DR EMBL; M26326; AAAS9463.1; -.
DR EMBL; X12876; CAA31369.1; -.
DR EMBL; X12881; CAA31375.1; -.
DR EMBL; X12883; CAA31377.1; -.
DR EMBL; M26325; AAAS9462.1; -.
DR PIR; S05481; S05481.
DR GLYCOSUITE; P05783; -.
DR SWISS-2DPAGE; P05783; HUMAN.
DR PMMA-2DPAGE; P05783; -.
DR SIENA-2DPAGE; P05783; -.
DR GENEW; HGNC:6430; KRT18.
DR MIM; 148070; -.
DR GO; GO:0005882; C:intermediate filament; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR INTERPRO; IPR001664; IF.
DR INTERPRO; IPR002957; Keratin_I.
DR PRAM; PR00038; filament; I.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
DR KW Intermediate filament; Coiled coil; Keratin; Glycoprotein;
KW Acetylation; Disease mutation.
FT INIT_MET 0 0
FT MOD_RES 0 1
FT DOMAIN 1 78 ACETYLATION.
FT DOMAIN 79 386 HEAD.
FT DOMAIN 387 429 ROD.
FT DOMAIN 79 114 TAIL.
FT DOMAIN 79 114 COIL 1A.
FT DOMAIN 115 131 LINKER 1.
FT DOMAIN 132 223 LINKER 1B.
FT DOMAIN 224 247 COIL 2.
FT DOMAIN 248 386 COIL 2.
FT SITE 270 STUTTER.
FT SITE 330 STUTTER.
FT CARBOHYD 29 29 O-LINKED (GLCNAC).
FT CARBOHYD 30 30 /FTID-CAR_000175.
FT CARBOHYD 48 48 O-LINKED (GLCNAC).
FT CARBOHYD 48 48 /FTID-CAR_000193.
FT VARIANT 127 127 H -> L (IN CC; INTERFERS WITH THE ABILITY
FT TO FORM NORMAL FILAMENTS).
FT CONFLICT 167 167 /FTID=VAR_003852.
FT CONFLICT 201 201 Y -> H (IN REF. 2; AAH00698).
FT CONFLICT 245 245 E -> Q (IN REF. 4).
FT CONFLICT 308 308 A -> S (IN REF. 4).
FT CONFLICT 311 311 D -> R (IN REF. 4).
FT CONFLICT 311 311 S -> R (IN REF. 4).
SQ SEQUENCE 429 AA; 47926 MW; 717E9D6A3933F8 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDAID 78
DB 390 NLGDAID 396
|||||

RESULT 36
SYTL_CABEL STANDARD; PRT; 441 AA.
AC F34693;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Synaptotagmin I.
GN SNT-1 OR F31E8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

```

OC NCBI_TaxID=6239;
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=93313960; PubMed=8391930;
RA Nonet M.L., Grundahl K., Meyer B.J., Rand J.B.;
RT "Synaptic function is impaired but not eliminated in C. elegans
RL mutants lacking synaptotagmin.";
RL Cell 73:1291-1305(1993).
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY HAVE A REGULATORY ROLE IN THE MEMBRANE INTERACTIONS
CC DURING TRAFFICKING OF SYNAPTIC VESICLES AT THE ACTIVE ZONE OF THE
CC SYNAPSE. IT BINDS ACIDIC PHOSPHOLIPIDS WITH A SPECIFICITY THAT
CC REQUIRES THE PRESENCE OF BOTH AN ACIDIC HEAD GROUP AND A DIACYL
CC BACKBONE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLES AND VESICLE-LIKE
CC STRUCTURES.
CC -!- TISSUE SPECIFICITY: LOCALIZED TO REGIONS KNOWN TO BE RICH IN
CC SYNAPSES AND APPEARS TO BE ASSOCIATED WITH SYNAPTIC VESICLES.
CC ALSO FOUND IN SOME NONNEURONAL SECRETORY STRUCTURES.
CC -!- DISEASE: MUTANTS EXHIBIT SEVERE BEHAVIORAL ABNORMALITIES THAT ARE
CC CHARACTERISTIC OF DEFICIENCIES IN SYNAPTIC FUNCTION, INCLUDING
CC SEVERE LOCOMOTION, FEEDING, AND DEFECATION DEFECTS.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; LI5302; AAA28145.1; -.
DR EMBL; U55856; AAA98023.1; -.
DR PIR; A40707; A40707.
DR HSPP; P21707; IRSY.
DR WormPep; F31E8.2; CE02711.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; IRI.
DR Pfam; PR00168; C2; 2.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMIN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR KW Transmembrane; Repeat; Synapse; Glycoprotein.
FT DOMAIN 1 69 VESICULAR (POTENTIAL).
FT TRANSMEM 70 96 POTENTIAL.
FT DOMAIN 97 441 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 175 262 C2 DOMAIN 1.
FT DOMAIN 306 397 C2 DOMAIN 2.
SQ SEQUENCE 441 AA; 49904 MW; F8D174337EB472DB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 441;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 TSEAVKQ 418
DB 32 TSEAVKQ 38
|||||

RESULT 37
CD4_CANFA STANDARD; PRT; 463 AA.
ID_CD4_CANFA

```


AC P33705;
 DT 01-FEB-1994 (Rel. 28, Last created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
 DE T4/Leu-3).
 GN CD4.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE OF 13-463 FROM N.A.
 RC STRAIN=Beagle; TISSUE=Thymus;
 RX MEDLINE=93192324; PubMed=7916632;
 RA Milde K.F., Conner G.E., Minz D.H., Alejandro R.;
 RT "Primary structure of the canine CD4 antigen."
 RL Biochim. Biophys. Acta 1172:315-318 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Beagle; TISSUE=Thymus;
 RX MEDLINE=94378217; PubMed=8091416;
 RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
 RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
 RT alpha antigens."
 RL Tissue Antigens 43:184-188 (1994).
 CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
 CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION.
 CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MACROPHAGES AND A SUBSET OF
 CC T LYMPHOCYTES.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L06130; AAB02295.1; -;
 DR EMBL; X68565; -; NOT_ANNOTATED_CDS.
 DR HSP; P01730; 1WBR.
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding activity; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
 DR GO; GO:0030217; P:T-cell differentiation; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
 DR InterPro; IPR000973; CD4 TCAG.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig.MHC.
 DR Pfam; PF00047; Ig; 3.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS00835; IG LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 24
 FT CHAIN 25 463 T-CELL SURFACE GLYCOPROTEIN CD4.
 FT DOMAIN 25 401 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 402 423 POTENTIAL.
 FT DOMAIN 424 463 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 124 IG-LIKE V-TYPE.
 FT DOMAIN 125 211 IG-LIKE C2-TYPE 1.
 FT DOMAIN 212 321 IG-LIKE C2-TYPE 2.
 FT DOMAIN 322 378 IG-LIKE C2-TYPE 3.

FT DISULFID 41 109 BY SIMILARITY.
 FT DISULFID 332 374 BY SIMILARITY.
 FT LIPID 424 424 PALMITATE (BY SIMILARITY).
 FT LIPID 427 427 PALMITATE (BY SIMILARITY).
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 463 AA; 51639 MW; 95805170CB4A833 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 463;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 365 KAGDAVE 371
 Db 32 KAGDAVE 38
 RESULT 38
 SG3_MOUSE
 ID SG3_MOUSE STANDARD; PRT; 471 AA.
 AC P47867;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Secretogranin III precursor (SgIII).
 GN SG3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95001263; PubMed=7917832;
 RA Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
 RA Sutcliffe J.G.;
 RT "Primary structure of mouse secretogranin III and its absence from
 RT deficient mice."
 R J. Mol. Neurosci. 4:225-233 (1993).
 CC -1- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
 CC secretory granules.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U02982; AAA56636.1; -;
 DR MGD; MGI:103032; Scg3.
 KW Signal; Cleavage on pair of basic residues.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 471 SECRETOGRANIN III.
 SQ SEQUENCE 471 AA; 53326 MW; FC1E9C381AFCA564 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 471;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 160 DGKIPEK 166
 Db 226 DGKIPEK 232
 RESULT 39
 SG3_RAT
 ID SG3_RAT STANDARD; PRT; 471 AA.

AC P47968;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Secretogranin III precursor (SgIII) (1B1075).
GN SCG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP MEDLINE=95001263; PubMed=7917832;
RX Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA Sutcliffe J.G.;
RT "Primary structure of mouse secretogranin III and its absence from
RT deficient mice."
RL J. Mol. Neurosci. 4:225-233(1993).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=90376160; PubMed=2204688;
RA Ottiger H.-P., Battenberg E.F., Tsou A.-P., Bloom F.E.,
RA Sutcliffe J.G.;
RT "1B1075: a brain- and pituitary-specific mRNA that encodes a novel
RT chromogranin / secretogranin-like component of intracellular
RT vesicles."
RL J. Neurosci. 10:3135-3147(1990).
CC -1- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC -1- TISSUE SPECIFICITY: BRAIN- AND PITUITARY-SPECIFIC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U02983; AAA56637.1; -
KW Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 471 SECRETGRANIN III.
SQ SEQUENCE 471 AA; 53183 MW; 977B3F8885F33223 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 DGKIPEK 166
Db 226 DGKIPEK 232
RESULT 40
Y213 MYCCE
ID Y213 MYCCE STANDARD; PRT; 471 AA.
AC P47455;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG213.
GN MG213.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.P., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudak D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -1- SIMILARITY: SOME, TO B.SUBTILIS YPUG.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U39701; AAC71432.1; -
DR PIR; E64223; E64223.
DR TIGR; MG213; -
DR InterPro; IPR003768; DUF173.
DR Pfam; PF02616; DUF173; 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 422 427 POLY-GLN.
SQ SEQUENCE 471 AA; 54829 MW; 9531DEC11D790FE2 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 724 KADNLLN 730
Db 217 KADNLLN 223
Search completed: November 14, 2003, 11:07:16
Job time : 28 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:05:49 ; Search time 21 Seconds
(without alignments)
3471.231 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 758
Sequence: 1 MAQTILKPIVLILLINTPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR 76:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	23.1	764	2 D81893	probable outer-mem
2	9	1.2	205	2 A72479	hypothetical prote
3	9	1.2	323	2 T05478	peroxidase (EC 1.1
4	9	1.2	1980	2 S54307	myosin heavy chain
5	9	1.2	2022	2 A59256	myosin-IxB [simila
6	8	1.1	99	2 AB1484	hypothetical prote
7	8	1.1	99	2 AH1123	hypothetical prote
8	8	1.1	196	2 AE2883	nitroreductase [l
9	8	1.1	196	2 C97659	probable enzyme (A
10	8	1.1	316	1 PNB5LC	beta-lactamase (EC
11	8	1.1	393	2 JU0334	formate dehydrogen
12	8	1.1	399	2 T19180	hypothetical prote
13	8	1.1	400	2 JC7815	formate dehydrogen
14	8	1.1	401	2 C95233	probable NAD-depen
15	8	1.1	544	2 S76859	hypothetical prote
16	8	1.1	892	2 T40040	Grpase-activator p
17	8	1.1	1009	2 S61174	hypothetical prote
18	7	0.9	54	2 S34567	gene L protein - h
19	7	0.9	75	2 AB1783	hypothetical prote
20	7	0.9	100	2 E89460	conserved hypothet
21	7	0.9	108	2 A72224	hypothetical prote
22	7	0.9	117	2 S01185	NADH2 dehydrogenas
23	7	0.9	117	2 G25797	hypothetical prote
24	7	0.9	117	2 C87588	hypothetical prote
25	7	0.9	125	2 T05445	hypothetical prote
26	7	0.9	141	2 T27080	hypothetical prote
27	7	0.9	141	2 T27071	hypothetical prote
28	7	0.9	149	2 T35846	probable integral
29	7	0.9	153	1 DTECR	aspartate carbamoy

30	7	0.9	153	1 DTEBCT	aspartate carbamoy
31	7	0.9	153	2 E91281	aspartate carbamoy
32	7	0.9	153	2 F86122	aspartate carbamoy
33	7	0.9	153	2 AC1059	hypothetical prote
34	7	0.9	155	2 S76912	hypothetical prote
35	7	0.9	166	2 T07089	dehydrin - soybean
36	7	0.9	171	2 AF2864	acetyltransferase
37	7	0.9	171	2 E97841	hypothetical prote
38	7	0.9	173	2 C70228	conserved hypothet
39	7	0.9	185	2 A87609	hypothetical prote
40	7	0.9	200	2 A41740	H+-transporting tw
41	7	0.9	201	2 F72285	transcription regu
42	7	0.9	202	2 F72739	hypothetical prote
43	7	0.9	212	2 B84337	deoxyribose-phosph
44	7	0.9	213	2 T33566	hypothetical prote
45	7	0.9	225	2 D64314	conserved hypothet
46	7	0.9	230	2 I37095	gene 2.19 protein
47	7	0.9	234	2 A90402	flagella-related p
48	7	0.9	241	2 G89925	menaquinone biosyn
49	7	0.9	244	2 H69844	diadenosine tetrap
50	7	0.9	249	2 D87533	hypothetical prote
51	7	0.9	249	2 S09868	hypothetical prote
52	7	0.9	249	2 E72646	probable molybdopt
53	7	0.9	250	2 B69038	conserved hypothet
54	7	0.9	251	2 AG0132	probable hydroxyac
55	7	0.9	251	2 T32200	hypothetical prote
56	7	0.9	263	2 A72349	conserved hypothet
57	7	0.9	265	2 T50569	probable membrane
58	7	0.9	278	2 T33129	hypothetical prote
59	7	0.9	280	1 D69513	conserved hypothet
60	7	0.9	285	2 A84386	hypothetical prote
61	7	0.9	294	2 T26946	hypothetical prote
62	7	0.9	295	2 B82674	conserved hypothet
63	7	0.9	303	2 B88241	protein T22C8.1 [l
64	7	0.9	303	2 T25114	hypothetical prote
65	7	0.9	310	2 C83909	oligopeptide ABC t
66	7	0.9	312	2 F87126	probable short cha
67	7	0.9	312	2 G82105	flagellar protein
68	7	0.9	315	2 A84598	probable triosepho
69	7	0.9	316	2 E81321	probable cation tr
70	7	0.9	320	2 T33662	hypothetical prote
71	7	0.9	325	2 T7553	brain mitochondria
72	7	0.9	327	2 T27483	hypothetical prote
73	7	0.9	333	1 JS0590	endo-1,4-beta-xyla
74	7	0.9	335	2 T50601	endo-1,4-beta-xyla
75	7	0.9	338	2 F84262	hypothetical prote
76	7	0.9	343	2 T26594	hypothetical prote
77	7	0.9	349	2 G69351	N2, N2-dimethylguan
78	7	0.9	353	2 T48310	hypothetical prote
79	7	0.9	368	2 A82093	membrane-bound lyl
80	7	0.9	369	2 JQ2278	hydroxymethylbilan
81	7	0.9	388	1 S72995	alanine racemase (
82	7	0.9	389	2 A38302	pepsin (EC 3.4.23.
83	7	0.9	391	2 AD2501	hypothetical prote
84	7	0.9	394	2 B70206	hypothetical prote
85	7	0.9	401	2 T18936	phosphoprotein pho
86	7	0.9	403	2 T18936	General secretion
87	7	0.9	406	2 H70364	iron-sulfur cofact
88	7	0.9	414	2 A12823	aminotransferase (
89	7	0.9	414	2 H97601	hypothetical prote
90	7	0.9	415	2 C70552	hypothetical prote
91	7	0.9	430	2 S05481	keratin 18, type I
92	7	0.9	432	2 S30193	T-cell surface gly
93	7	0.9	434	2 D72353	lipopolysaccharide
94	7	0.9	437	2 T34211	hypothetical prote
95	7	0.9	441	2 A40707	synaptotagmin - Ca
96	7	0.9	442	2 C96688	unknown protein, 2
97	7	0.9	443	2 B86180	protein T1G11.3 [i
98	7	0.9	444	2 JC4348	virulence-mediatin
99	7	0.9	448	2 T01814	hypothetical prote
100	7	0.9	468	2 E81924	probable two-compo
101	7	0.9	468	2 G81180	sensor histidine k
102	7	0.9	468	2 T22287	hypothetical prote

103	7	0.9	469	2	D82976	hypothenical prote	176	7	0.9	2142	2	D86303	F17F16.1 protein -
104	7	0.9	469	2	T48783	related to SGT1 pr	177	7	0.9	3263	2	D82410	hypothenical prote
105	7	0.9	469	2	E86421	hypothenical prote	178	7	0.9	3433	1	S28381	utrophin - human
106	7	0.9	470	2	AF3435	tlldD protein limpo	179	7	0.9	4302	2	A38971	polycystic kidney
107	7	0.9	471	2	E64223	hypothenical prote	180	7	0.9	4485	2	T08044	dynein gamma heavy
108	7	0.9	488	2	A81606	glycine dehydrogen	181	6	0.8	20	2	A49142	hyaluronidase (EC
109	7	0.9	488	2	AF1243	glycine dehydrogen	182	6	0.8	22	2	T09010	N4-(beta-N-acetyl)g
110	7	0.9	490	2	AE3433	mitochondrial proc	183	6	0.8	27	2	B43853	probable phosphogl
111	7	0.9	492	2	T02458	hypothenical prote	184	6	0.8	33	2	T08018	ycf12 protein - Ch
112	7	0.9	497	2	T48676	proline uptake pro	185	6	0.8	50	2	AG2783	hypothenical prote
113	7	0.9	498	2	AD7374	hypothenical prote	186	6	0.8	50	2	G97562	hypothenical prote
114	7	0.9	509	2	D86794	unknown protein F1	187	6	0.8	51	2	PN0479	cell division cycl
115	7	0.9	512	2	E89978	high affinity prol	188	6	0.8	57	2	H95181	hypothenical prote
116	7	0.9	513	2	T37180	probable membrane	189	6	0.8	57	2	C98049	hypothenical prote
117	7	0.9	524	2	D70861	probable monooxygen	190	6	0.8	59	2	E84011	probable regulator
118	7	0.9	533	2	T13607	hypothenical prote	191	6	0.8	60	2	A10343	hypothenical prote
119	7	0.9	533	2	A37180	chromogranin/secre	192	6	0.8	62	2	G64561	hypothenical prote
120	7	0.9	537	2	B84772	probable Dnad prot	193	6	0.8	62	2	D89907	hypothenical prote
121	7	0.9	537	2	G31277	guinate transport	194	6	0.8	64	2	F86693	hypothenical prote
122	7	0.9	539	2	S58287	J-domain protein D	195	6	0.8	66	2	B99970	hypothenical prote
123	7	0.9	545	2	S46151	probable purine nu	196	6	0.8	70	2	T44130	hypothenical prote
124	7	0.9	555	2	H83043	hypothenical prote	197	6	0.8	71	2	AF2940	hypothenical prote
125	7	0.9	558	2	D70449	conserved hypotet	198	6	0.8	71	2	C98342	hypothenical prote
126	7	0.9	569	2	T50711	urease (EC 3.5.1.5	199	6	0.8	72	2	A12928	conserved hypotet
127	7	0.9	594	2	E96667	unknown protein, 6	200	6	0.8	72	2	E98353	hypothenical prote
128	7	0.9	601	2	B75136	hypothenical prote	201	6	0.8	73	2	B83338	hypothenical prote
129	7	0.9	608	2	A35021	myosin-light-chain	202	6	0.8	73	2	F83266	hypothenical prote
130	7	0.9	623	2	H87127	conserved hypotet	203	6	0.8	75	2	A13219	hypothenical prote
131	7	0.9	628	2	A87596	hypothenical prote	204	6	0.8	82	2	F95086	conserved hypotet
132	7	0.9	642	2	E87644	sensory box histid	205	6	0.8	82	2	A97954	conserved hypotet
133	7	0.9	645	2	D85680	unknown protein en	206	6	0.8	85	2	S14026	hypothenical prote
134	7	0.9	645	2	D85844	major head protein	207	6	0.8	85	2	C82589	hypothenical prote
135	7	0.9	645	2	A90853	major head protein	208	6	0.8	87	2	T09816	26S proteinase reg
136	7	0.9	645	2	G90821	major head protein	209	6	0.8	87	2	T14135	NADH2 dehydrogenas
137	7	0.9	645	2	G90821	major head protein	210	6	0.8	89	2	T09819	26S proteinase reg
138	7	0.9	645	2	D90910	major head protein	211	6	0.8	89	2	C72286	transcription regu
139	7	0.9	645	2	D90875	major head protein	212	6	0.8	92	2	PQ0632	coat protein - Rem
140	7	0.9	648	2	S59723	transcription acti	213	6	0.8	94	2	D82498	conserved hypotet
141	7	0.9	658	2	C75187	anaerobic ribonucl	214	6	0.8	95	2	A03858	hypothenical prote
142	7	0.9	676	2	S70395	env polyprotein -	215	6	0.8	100	2	T38899	hypothenical prote
143	7	0.9	720	2	S69890	mitogen inducible	216	6	0.8	100	2	H81310	probable periplasm
144	7	0.9	728	2	T27884	hypothenical prote	217	6	0.8	100	2	AH2220	hypothenical prote
145	7	0.9	737	2	AD2037	hypothenical prote	218	6	0.8	100	2	A10092	probable membrane
146	7	0.9	741	2	T13791	NADH2 dehydrogenas	219	6	0.8	101	1	W7ML51	E7 protein - human
147	7	0.9	745	2	C88618	protein ZK520.4 [i	220	6	0.8	101	2	JX0333	ribonuclease (EC 3
148	7	0.9	764	2	H83055	probable outer mem	221	6	0.8	101	2	A43262	hypothenical prote
149	7	0.9	765	2	AF2200	phytochrome A, two	222	6	0.8	101	2	E83252	hypothenical prote
150	7	0.9	765	2	T15447	hypothenical prote	223	6	0.8	103	2	I79265	hypothenical prote
151	7	0.9	770	2	T26783	hypothenical prote	224	6	0.8	104	1	ASLJ5Y	traa protein - Esc
152	7	0.9	782	2	E90427	hypothenical prote	225	6	0.8	104	2	T03316	vpr protein - huma
153	7	0.9	802	2	AF1227	phenylalanyl-trna	226	6	0.8	104	2	D95232	gene 19 protein -
154	7	0.9	802	2	AH1580	phenylalanyl-trna	227	6	0.8	104	2	D95232	hypothenical prote
155	7	0.9	818	2	F97772	hypothenical prote	228	6	0.8	107	2	B71069	hypothenical prote
156	7	0.9	844	1	RXSSSP	RNA-directed RNA p	229	6	0.8	107	2	B71074	hypothenical prote
157	7	0.9	847	2	T04772	hypothenical prote	230	6	0.8	108	2	F83886	hypothenical prote
158	7	0.9	881	2	A97107	alanyl-trna synthe	231	6	0.8	110	2	C97137	hypothenical prote
159	7	0.9	892	2	AG1661	transcription regu	232	6	0.8	112	2	C72679	hypothenical prote
160	7	0.9	927	2	A45039	CTF4 protein - yea	233	6	0.8	112	2	B84345	hypothenical prote
161	7	0.9	966	2	E84053	penicillin-binding	234	6	0.8	113	2	B83575	hypothenical prote
162	7	0.9	1007	2	D48535	rho-type GTPase-ac	235	6	0.8	115	2	A84676	60S acidic ribosom
163	7	0.9	1018	2	E83099	probable RND efflu	236	6	0.8	115	2	G72466	hypothenical prote
164	7	0.9	1044	2	T37568	hypothenical prote	237	6	0.8	116	2	S14211	NADH2 dehydrogenas
165	7	0.9	1046	2	H87318	TonB-dependent rec	238	6	0.8	116	2	S59099	ribosomal protein
166	7	0.9	1052	1	B49219	toxin iii - Actino	239	6	0.8	116	2	C69071	hypothenical prote
167	7	0.9	1067	2	S33417	kinesin-like prote	240	6	0.8	117	2	C69053	ribosomal protein
168	7	0.9	1086	2	T33893	hypothenical prote	241	6	0.8	117	2	F82308	hypothenical prote
169	7	0.9	1186	2	AG1928	two-component hybr	242	6	0.8	118	2	JE0398	ribosomal protein
170	7	0.9	1252	2	H97178	ATP-dependent exon	243	6	0.8	118	2	S70089	korA protein - Amy
171	7	0.9	1366	2	T35985	probable large Pro	244	6	0.8	119	1	YQECB2	fimbrial protein p
172	7	0.9	1400	2	T52359	hypothenical prote	245	6	0.8	119	1	YQECR1	fimbrial protein p
173	7	0.9	1786	1	H71527	probable excinucle	246	6	0.8	119	2	S03302	T-cell receptor ga
174	7	0.9	1832	2	T31113	mucin-like glycopr	247	6	0.8	119	2	I40630	prepilin - Escheri
175	7	0.9	2089	1	A48757	acetyl-CoA carboxy	248	6	0.8	119	2	C70140	hypothenical prote

249	6	0.8	120	1	YQECR9	fimbrial protein p	322	6	0.8	149	2	F83964	hypothetical prote
250	6	0.8	120	2	T12918	hypothetical prote	323	6	0.8	150	2	C97826	hypothetical prote
251	6	0.8	121	1	RIAFSI	somatostatin I pre	324	6	0.8	150	2	E85514	unknown protein fr
252	6	0.8	121	1	YQECF	fimbrial protein p	325	6	0.8	150	2	H97158	hypothetical prote
253	6	0.8	121	1	YQECF	fimbrial protein p	326	6	0.8	151	2	T43478	hypothetical prote
254	6	0.8	121	1	B2AG58	virB2 protein prec	327	6	0.8	152	2	C81189	probable periplasm
255	6	0.8	121	2	AF3248	hypothetical prote	328	6	0.8	152	2	F72646	hypothetical prote
256	6	0.8	121	2	D72553	hypothetical prote	329	6	0.8	152	2	AI3271	acetyltransferase
257	6	0.8	121	2	D82604	hypothetical prote	330	6	0.8	152	2	D98270	hypothetical prote
258	6	0.8	122	2	C75317	hypothetical prote	331	6	0.8	153	2	S08501	3-dehydroquinat d
259	6	0.8	122	2	T01558	auxin-induced prot	332	6	0.8	153	2	AF2203	bacterioferritin c
260	6	0.8	122	2	S48325	probable membrane	333	6	0.8	153	2	AF2344	hypothetical prote
261	6	0.8	123	1	RHPCN	galanin precursor	334	6	0.8	153	2	T31654	hypothetical prote
262	6	0.8	124	2	A83505	hypothetical prote	335	6	0.8	154	1	XMECF1	flagellar protein
263	6	0.8	124	2	S75364	hypothetical prote	336	6	0.8	154	2	C90964	flagellar protein
264	6	0.8	124	2	S64773	probable membrane	337	6	0.8	154	2	C85812	flagellar biosynth
265	6	0.8	125	2	A96931	nitrogen regulator	338	6	0.8	154	2	E70902	probable riboflavi
266	6	0.8	126	2	S74493	thioredoxin M-2 -	339	6	0.8	154	2	E90794	urease accessory p
267	6	0.8	126	2	C33989	Ig heavy chain V-3	340	6	0.8	154	2	G85603	probable urease ac
268	6	0.8	128	2	A97248	nudix (MudT) fami	341	6	0.8	155	1	WMBV7B	17K protein - barl
269	6	0.8	128	2	D70742	hypothetical prote	342	6	0.8	155	2	A82250	phosphotyrosine pr
270	6	0.8	129	2	F95865	probable ribulose-	343	6	0.8	155	2	S33509	interleukin-2 - Mo
271	6	0.8	129	2	S40332	Ig kappa chain - h	344	6	0.8	155	2	B30929	flagellar protein
272	6	0.8	129	2	C75328	hypothetical prote	345	6	0.8	155	2	AH0752	FlaL protein (impo
273	6	0.8	130	2	A24574	hypothetical prote	346	6	0.8	155	2	A64801	yhey protein - Esc
274	6	0.8	130	2	D89817	T-cell receptor ga	347	6	0.8	155	2	A90716	hypothetical prote
275	6	0.8	130	2	C70971	hypothetical prote	348	6	0.8	155	2	D85566	hypothetical prote
276	6	0.8	130	2	C88102	protein W09G10.6 l	349	6	0.8	155	2	T34169	hypothetical prote
277	6	0.8	130	2	AC2314	hypothetical prote	350	6	0.8	156	2	AQ0754	patch repair prote
278	6	0.8	131	2	H83156	hypothetical prote	351	6	0.8	157	2	E72332	thioredoxin peroxi
279	6	0.8	131	2	C83863	hypothetical prote	352	6	0.8	158	2	E86993	hypothetical prote
280	6	0.8	132	2	D72616	hypothetical prote	353	6	0.8	158	2	D36138	ureE protein - Kle
281	6	0.8	133	2	B82977	conserved hypothet	354	6	0.8	158	2	H83824	hypothetical prote
282	6	0.8	134	2	D96568	probable histone H	355	6	0.8	159	2	B84669	hypothetical prote
283	6	0.8	134	2	G82411	conserved hypothet	356	6	0.8	160	2	T24339	hypothetical prote
284	6	0.8	134	2	AG2552	hypothetical prote	357	6	0.8	160	2	H90985	GBP-mannose mannos
285	6	0.8	134	2	D95296	hypothetical prote	358	6	0.8	160	2	C85831	GNP-mannose mannos
286	6	0.8	136	2	S12847	T-cell receptor ga	359	6	0.8	160	2	E55239	Colanic acid biosy
287	6	0.8	136	2	T47615	histone H2A.F/Z -	360	6	0.8	160	2	A81075	FRBP-type peptidyl
288	6	0.8	136	2	E36790	hypothetical prote	361	6	0.8	161	2	T06785	nonsymbiotic hemog
289	6	0.8	136	2	F72353	flagellar protein	362	6	0.8	161	2	JQ2160	coat protein - tom
290	6	0.8	136	2	S76416	hypothetical prote	363	6	0.8	161	2	JQ2146	coat protein - tom
291	6	0.8	137	2	E69359	hypothetical prote	364	6	0.8	161	2	F64038	hypothetical prote
292	6	0.8	138	2	A37330	venom allergen II	365	6	0.8	161	2	C65170	pcs system, arbuti
293	6	0.8	139	2	B69386	conserved hypothet	366	6	0.8	162	2	D82983	leucine-responsive
294	6	0.8	139	2	A83579	conserved hypothet	367	6	0.8	162	2	T13522	hypothetical prote
295	6	0.8	139	2	H90298	hypothetical prote	368	6	0.8	162	2	D98020	conserved hypothet
296	6	0.8	139	2	G97240	amino acid transpo	369	6	0.8	162	2	B70144	transcription fact
297	6	0.8	140	2	T04904	hypothetical prote	370	6	0.8	163	2	B70854	hypothetical prote
298	6	0.8	141	2	JT0624	hemoglobin alpha 1	371	6	0.8	163	2	T10220	hypothetical prote
299	6	0.8	141	2	F69948	phage-related prot	372	6	0.8	164	2	B82888	single-stranded DN
300	6	0.8	141	2	AE3275	hypothetical cytos	373	6	0.8	164	2	S11846	rab16B protein - r
301	6	0.8	143	2	S13238	tail fiber assembl	374	6	0.8	164	2	S11847	rab16C protein - r
302	6	0.8	145	1	ZWECAP	muCA protein - Esc	375	6	0.8	164	2	S77233	riboflavin synthas
303	6	0.8	145	2	F83834	hypothetical prote	376	6	0.8	164	2	I40526	hypothetical prote
304	6	0.8	145	2	C72668	hypothetical prote	377	6	0.8	164	2	H89784	conserved hypothet
305	6	0.8	146	2	S06775	muCA protein - Sal	378	6	0.8	165	2	T15780	hypothetical prote
306	6	0.8	146	2	C83805	hypothetical prote	379	6	0.8	165	2	S34270	hypothetical prote
307	6	0.8	146	2	B64472	hypothetical prote	380	6	0.8	165	2	S34269	fibrinogen-binding
308	6	0.8	146	2	T10511	hypothetical prote	381	6	0.8	166	2	G75449	MutT/nudix family
309	6	0.8	146	2	AD0929	conserved hypothet	382	6	0.8	167	2	T36325	hypothetical prote
310	6	0.8	146	2	G65008	hypothetical prote	383	6	0.8	167	2	D55224	hypothetical prote
311	6	0.8	147	2	A29910	myosin calcium-bin	384	6	0.8	167	2	S34363	hypothetical prote
312	6	0.8	147	2	C64514	hypothetical prote	385	6	0.8	168	2	H69940	phosphotransferase
313	6	0.8	147	2	G75629	conserved hypothet	386	6	0.8	168	2	S72898	hypothetical prote
314	6	0.8	147	2	H70630	hypothetical prote	387	6	0.8	168	2	T45305	hypothetical prote
315	6	0.8	147	2	B90664	hypothetical prote	388	6	0.8	169	2	E26395	T-cell receptor be
316	6	0.8	148	2	T52550	probable transcrip	389	6	0.8	170	2	H82746	polypeptide deform
317	6	0.8	148	2	E95886	probable sugar-pho	390	6	0.8	171	2	S75677	hypothetical prote
318	6	0.8	148	2	H64004	hypothetical prote	391	6	0.8	171	2	AI1060	probable acetyltra
319	6	0.8	148	2	F95994	hypothetical prote	392	6	0.8	172	2	T27505	hypothetical prote
320	6	0.8	149	2	C97180	uncharacterized pr	393	6	0.8	173	2	H95961	hypothetical prote
321	6	0.8	149	2	D82178	hypothetical prote	394	6	0.8	174	1	A31096	fimbrial protein p

395	6	0.8	174	2	AF3297	biopolymer transpo	468	6	0.8	201	2	H82935	recombinational DN
396	6	0.8	174	2	T39690	probable actin pol	469	6	0.8	201	2	C88473	protein F40H6.3 [i
397	6	0.8	175	2	F37802	criz protein - Erv	470	6	0.8	201	2	R83319	hypothetical prote
398	6	0.8	175	2	AC3014	MutT/nudix family	471	6	0.8	202	1	MNIV47	nonstructural prot
399	6	0.8	175	2	G84383	hypothetical prote	472	6	0.8	202	2	E75209	hypothetical prote
400	6	0.8	176	2	F82478	hypothetical prote	473	6	0.8	202	2	T43982	glycoprotein [impo
401	6	0.8	176	2	B70445	heat shock protein	474	6	0.8	202	2	T44168	hypothetical prote
402	6	0.8	178	2	D69467	probable orotate p	475	6	0.8	202	2	G86302	Fl7Fl6.6 protein -
403	6	0.8	178	2	T20317	hypothetical prote	476	6	0.8	203	2	S51495	GTP-binding protei
404	6	0.8	179	2	G89817	hypothetical prote	477	6	0.8	203	2	F75554	MutT/nudix family
405	6	0.8	179	2	AT2056	polypeptide deform	478	6	0.8	204	2	H96027	probable protocate
406	6	0.8	179	2	T40442	hypothetical highl	479	6	0.8	204	2	D64324	hypothetical prote
407	6	0.8	180	2	S32917	hypothetical prote	480	6	0.8	204	2	C97866	hypothetical prote
408	6	0.8	180	2	H81181	bacteriocin resist	481	6	0.8	205	2	AE3600	hypothetical prote
409	6	0.8	181	2	T24914	hypothetical prote	482	6	0.8	205	2	AB3432	hypothetical prote
410	6	0.8	181	2	T33410	hypothetical prote	483	6	0.8	206	2	AG3114	2-haloacid dehalog
411	6	0.8	181	2	AC3915	hypothetical prote	484	6	0.8	206	2	F98172	protocatechuate 3,
412	6	0.8	181	2	F89927	conserved hypothet	485	6	0.8	207	2	S54128	hypothetical 21.4K
413	6	0.8	182	2	F81366	adenine phosphorib	486	6	0.8	207	2	AB3016	hypothetical prote
414	6	0.8	182	2	C87097	conserved hypothet	487	6	0.8	207	2	C90534	hypothetical prote
415	6	0.8	182	2	S78899	hypothetical prote	488	6	0.8	207	2	AB1877	general secretion
416	6	0.8	182	2	F37985	hypothetical prote	489	6	0.8	208	2	S03615	glutathione transf
417	6	0.8	183	1	TUBP84	tail fiber assembl	490	6	0.8	208	2	I64209	ribosomal protein
418	6	0.8	183	2	S13240	tail fiber assembl	491	6	0.8	208	2	F87633	hypothetical prote
419	6	0.8	183	2	A89879	hypothetical prote	492	6	0.8	209	1	D35119	protocatechuate 3,
420	6	0.8	183	2	T07241	hypothetical prote	493	6	0.8	210	2	C75380	uracil phosphoribo
421	6	0.8	184	2	A61628	early gland protei	494	6	0.8	210	2	F82238	phosphoribosyl-AMP
422	6	0.8	184	2	B30592	hypothetical prote	495	6	0.8	210	2	F71308	probable 2-dehydro
423	6	0.8	185	2	C83644	conserved hypothet	496	6	0.8	210	2	T33697	hypothetical prote
424	6	0.8	187	2	H64995	hypothetical prote	497	6	0.8	210	2	C70528	hypothetical prote
425	6	0.8	187	2	B21021	hypothetical prote	498	6	0.8	210	2	A87331	hypothetical prote
426	6	0.8	187	2	A85865	hypothetical prote	499	6	0.8	210	2	T26154	hypothetical prote
427	6	0.8	187	2	A95896	conserved hypothet	500	6	0.8	210	2	A27710	hypothetical prote
428	6	0.8	188	2	AH0588	probable DNA recom	501	6	0.8	211	2	S73791	hypothetical prote
429	6	0.8	188	2	A84214	adenine phosphorib	502	6	0.8	211	2	T09316	glycoprotein - hum
430	6	0.8	189	2	A25556	hypothetical 21.2K	503	6	0.8	211	2	D97775	hypothetical prote
431	6	0.8	189	2	F96593	hypothetical prote	504	6	0.8	211	2	C96539	hypothetical prote
432	6	0.8	189	2	AG1947	hypothetical prote	505	6	0.8	212	1	F65034	L-fucose-phospha
433	6	0.8	190	2	S70699	X4-transporing AT	506	6	0.8	212	1	B64166	cytochrome c bioge
434	6	0.8	190	2	A82305	6,7-dimethyl-8-rib	507	6	0.8	212	2	AG0854	probable sugar ald
435	6	0.8	191	2	T31903	hypothetical prote	508	6	0.8	212	2	A71702	hypothetical prote
436	6	0.8	191	2	T21031	hypothetical prote	509	6	0.8	212	2	B69493	hypothetical prote
437	6	0.8	192	2	T17189	conserved hypothet	510	6	0.8	214	2	B95898	probable sensory t
438	6	0.8	192	2	A82277	tellurite resistan	511	6	0.8	214	2	T33352	hypothetical prote
439	6	0.8	192	2	G87452	hypothetical prote	512	6	0.8	214	2	T22487	hypothetical prote
440	6	0.8	192	2	A86398	protein T7N9.22 [i	513	6	0.8	215	2	A49677	endoplasmic reticu
441	6	0.8	192	2	F82126	hypothetical prote	514	6	0.8	215	2	I51377	hypothetical prote
442	6	0.8	193	2	S67201	Maf/YceF/YndE fami	515	6	0.8	215	2	G70551	hypothetical prote
443	6	0.8	193	2	T96581	hypothetical prote	516	6	0.8	215	2	C86962	probable oligoribo
444	6	0.8	194	2	I57523	HSP90 - mouse (fra	517	6	0.8	215	2	A80793	conserved hypothet
445	6	0.8	194	2	D75428	hypothetical prote	518	6	0.8	216	2	JI0090	ependymin precurs
446	6	0.8	194	2	T19641	hypothetical prote	519	6	0.8	216	2	A32636	ependymin II precu
447	6	0.8	194	2	T19642	hypothetical prote	520	6	0.8	216	2	S64240	probable membrane
448	6	0.8	195	2	AD2243	hypothetical prote	521	6	0.8	216	2	H84241	hypothetical prote
449	6	0.8	195	2	T19643	hypothetical prote	522	6	0.8	216	2	H87587	hypothetical prote
450	6	0.8	196	2	AD0385	4-methyl-5(B-hyaro	523	6	0.8	216	2	AH0685	probable lipoprote
451	6	0.8	196	2	G72063	15 kDa Cysteine-ri	524	6	0.8	216	2	B87674	conserved hypothet
452	6	0.8	196	2	H86559	15 kDa Cysteine-ri	525	6	0.8	217	1	JS0630	formate dehydrogen
453	6	0.8	196	2	F90785	probable enzyme [i	526	6	0.8	217	1	MNIV61	nonstructural prot
454	6	0.8	196	2	D85645	probable enzyme Z1	527	6	0.8	217	2	A85729	formate dehydrogen
455	6	0.8	196	2	F64842	hypothetical prote	528	6	0.8	217	2	H90888	formate dehydrogen
456	6	0.8	197	2	S46928	phytochrome - Char	529	6	0.8	217	2	F82788	thymidylate kinase
457	6	0.8	197	2	A81912	probable GTP cyclo	530	6	0.8	217	2	F70645	hypothetical prote
458	6	0.8	197	2	G81104	GTP cyclohydrolase	531	6	0.8	217	2	T42053	redz protein - Str
459	6	0.8	197	2	A82556	c-type cytochrome	532	6	0.8	217	2	AF3050	two component resp
460	6	0.8	197	2	S35252	proline-rich prote	533	6	0.8	217	2	F98235	probable two-compo
461	6	0.8	197	2	T43066	hypothetical prote	534	6	0.8	217	2	A43820	ependymin precurs
462	6	0.8	198	2	H69487	SSU ribosomal prot	535	6	0.8	217	2	H70631	hypothetical prote
463	6	0.8	198	2	AB1391	probable transcrip	536	6	0.8	217	2	AF1871	hypothetical prote
464	6	0.8	199	2	T08902	manganese-binding	537	6	0.8	217	2	T46468	hypothetical prote
465	6	0.8	199	2	T36622	hypothetical prote	538	6	0.8	218	2	F83857	phosphoribosyl ant
466	6	0.8	200	2	H84715	probable phytocyan	539	6	0.8	218	2	C97236	ABC transporter, A
467	6	0.8	200	2	B81322	probable integral	540	6	0.8	218	2	S73675	hypothetical prote

541	6	0.8	218	2	T25229	hypothetical prote	614	1	MNIV77	nonstructural prot
542	6	0.8	218	2	T03383	zinc-induced prote	615	237	MNIVC1	nonstructural prote
543	6	0.8	219	2	D71001	hypothetical prote	616	237	F84687	hypothetical prote
544	6	0.8	219	2	T02112	probable bZIP tran	617	237	S58729	probable ribosomal
545	6	0.8	219	2	E83399	hypothetical prote	618	238	H72778	hypothetical prote
545	6	0.8	219	2	AC2135	hypothetical prote	619	238	G83528	hypothetical prote
547	6	0.8	220	2	I50588	fibroblast growth	620	239	E64699	hypothetical prote
548	6	0.8	220	2	T39548	hypothetical prote	621	239	T49360	related to a-agglu
549	6	0.8	221	2	C81099	hypothetical prote	622	240	WMNVP6	p26 protein - Auto
550	6	0.8	221	2	C81842	hypothetical prote	623	240	T24363	hypothetical prote
551	6	0.8	221	2	G22322	glutaredoxin-relat	624	240	T41874	p26 protein (simil
552	6	0.8	221	2	C64040	hypothetical prote	625	241	S47673	leucine transport
553	6	0.8	221	2	F71668	hypothetical prote	626	241	E86012	ATP-binding compon
554	6	0.8	222	2	G95266	probable ABC trans	627	241	E91166	ATP-binding compon
555	6	0.8	222	2	A75063	flagellin B precur	628	241	S53812	EMGARA beta isofo
556	6	0.8	222	2	H89796	conserved hypotet	629	241	T25886	hypothetical prote
557	6	0.8	222	2	C97822	hypothetical prote	630	241	F91287	hypothetical prote
558	6	0.8	223	2	G83498	hypothetical prote	631	241	A86129	hypothetical prote
559	6	0.8	223	2	T35665	hypothetical prote	632	241	S56536	hypothetical prote
560	6	0.8	223	2	T29976	hypothetical prote	633	241	A45724	pectate lyase [EC
561	6	0.8	224	2	A83569	ribulose-phosphate	634	242	A86189	protein T25820.7 [
562	6	0.8	224	2	AF3646	1-fucose phospho	635	242	B70570	hypothetical prote
563	6	0.8	224	2	S44964	lmbO protein - Str	636	243	F71861	hypothetical prote
564	6	0.8	224	2	AG1795	hypothetical prote	637	243	D64568	conserved hypotet
565	6	0.8	224	2	F84239	hypothetical prote	638	243	D90549	ribonuclease hii (
566	6	0.8	225	2	S76473	probable 1-acylgly	639	243	D83504	probable transcrip
567	6	0.8	225	2	A99206	hypothetical prote	640	243	S74031	hypothetical prote
568	6	0.8	225	2	AH3080	RntB family transp	641	243	AC3600	cellulase [EC 3.2.
569	6	0.8	225	2	AE2539	hypothetical prote	642	244	JC4708	gelatin-binding 28
570	6	0.8	226	2	C87637	conserved hypotet	643	244	T17916	hypothetical prote
571	6	0.8	226	2	H87291	conserved hypotet	644	245	JE0213	hisa protein - Cor
572	6	0.8	226	2	T43814	conserved hypotet	645	246	S47805	hypothetical 27.4K
573	6	0.8	227	1	MNIVX1	nonstructural prot	646	246	D91186	probable outer mem
574	6	0.8	227	1	MNIVX3	nonstructural prot	647	246	C86033	hypothetical prote
575	6	0.8	227	1	MNIVX5	nonstructural prot	648	246	E70861	C-terminal domain
576	6	0.8	227	1	MNIVX9	nonstructural prot	649	246	C97177	surface lipoprotei
577	6	0.8	227	1	MNIVX7	nonstructural prot	650	248	S34995	ankyrin 3 homolog
578	6	0.8	227	1	I54426	MHC H2-w28-E alpha	651	249	T04436	probable oxidoredu
579	6	0.8	227	2	E90432	triosephosphate is	652	250	B71859	short chain alcoho
580	6	0.8	228	2	D82874	tRNA (Guanine-N1) -	653	250	E64564	hypothetical prote
581	6	0.8	228	2	T22219	hypothetical prote	654	250	C83678	hypothetical prote
582	6	0.8	229	2	B69409	conserved hypotet	655	250	AG3276	precorrin-3 methyl
583	6	0.8	229	2	T33141	hypothetical prote	656	251	AG3587	SN-glycerol-3-phos
584	6	0.8	230	1	VG1HBC	E1 membrane glycop	657	251	T03535	precorrin-6x reduc
585	6	0.8	230	1	QJ1749	E1 membrane glycop	658	251	G83188	undecaprenyl pyrop
586	6	0.8	230	1	MNIV1	nonstructural prot	659	251	S41587	hemd protein - Pse
587	6	0.8	230	1	MNIV1F	nonstructural prot	660	251	A82989	uroporphyrinogen-I
588	6	0.8	230	1	MNIV1A1	nonstructural prot	661	251	T35258	probable transcrip
589	6	0.8	230	1	MNIV1A2	nonstructural prot	662	252	A75073	phosphoesterase-re
590	6	0.8	230	1	MNIV1A3	nonstructural prot	663	252	S69786	pgk protein - Sal
591	6	0.8	230	1	MNIV1A4	nonstructural prot	664	252	AC0849	pathogenicity 1 is
592	6	0.8	230	2	QJ1172	membrane protein -	665	252	H82574	phage-related prot
593	6	0.8	230	2	S09648	nonstructural prot	666	253	F71667	hypothetical prote
594	6	0.8	230	2	D97697	exu regulon regula	667	253	D97820	zinc/manganese ABC
595	6	0.8	230	2	A75260	hypothetical prote	668	253	T30928	hypothetical prote
596	6	0.8	230	2	AD2923	transcription regu	669	254	H69057	hypothetical prote
597	6	0.8	231	2	H84473	hypothetical prote	670	254	E90969	hypothetical prote
598	6	0.8	232	2	C70699	probable paba prot	671	254	G69495	transcription regu
599	6	0.8	232	2	T10008	probable p-aminobe	672	254	D85742	hypothetical prote
600	6	0.8	232	2	D71208	hypothetical prote	673	255	1 HLMSED	H-2 class II histo
601	6	0.8	232	2	T15146	hypothetical prote	674	255	1 HLMSEA	H-2 class II histo
602	6	0.8	232	2	A61045	homeotic protein T	675	255	S65032	H-2-transporing tw
603	6	0.8	232	2	B71325	probable V-type AT	676	255	D87355	RNA pseudouridylat
604	6	0.8	233	1	D69630	menaquinone biosyn	677	255	G64320	hypothetical prote
605	6	0.8	233	2	D75185	hypothetical prote	678	255	H86321	hypothetical prote
606	6	0.8	233	2	G83061	hypothetical prote	679	255	B55218	gramicidin S biosy
607	6	0.8	234	2	T20933	hypothetical prote	680	256	S58743	H+-transporing tw
608	6	0.8	234	2	H84297	phosphoglycerate d	681	256	AH2785	stationary-phase s
609	6	0.8	235	2	AG2218	hypothetical prote	682	256	E97565	stationary-phase s
610	6	0.8	236	2	E97723	hypothetical prote	683	256	G96774	hypothetical prote
611	6	0.8	236	2	A53853	apolipoprotein B m	684	256	T35290	probable transfera
612	6	0.8	236	2	D71548	hypothetical prote	685	256	F70812	probable lpqr prot
613	6	0.8	237	1	MNIV14	nonstructural prot	686	256	B82097	conserved hypotet

687 256 2 H82235 response regulator
688 1 S39747 ywfn protein - Bac
689 258 1 E84427 hypothetical prote
690 259 2 B84277 integral membrane
691 260 2 T18554 hypothetical prote
692 261 1 WNA087 early E1A 28K prot
693 262 2 T28643 conserved hypothet
694 263 2 T28643 Y4JP protein - Rhi
695 264 2 C95193 cell division prot
696 265 2 H87525 hydrolase, alpha/b
697 266 2 P83546 probable transcrip
698 267 2 A75076 membrane protein P
699 268 2 T47468 hypothetical prote
700 269 2 C83126 ferric enterobacti
701 270 2 D83577 conserved hypothet
702 271 2 B86338 protein F5M15.22 l
703 272 2 T24623 hypothetical prote
704 273 1 J50344 tryptophan synthas
705 274 2 H82387 glucosamine-6-phos
706 275 2 T52322 chlorophyll a/b-bi
707 276 2 G83874 oligopeptide ABC t
708 277 2 H98059 hypothetical prote
709 278 2 AC2383 hypothetical prote
710 279 2 C89818 dihydroperate sy
711 280 2 B92220 LSU ribosomal prot
712 281 2 H87469 ThiJ/FipI family p
713 282 2 A81280 probable sigma fac
714 283 1 A30584 interleukin-1 beta
715 284 2 T46749 hemolysin tlyC lva
716 285 2 D25564 probable myo-inosi
717 286 1 A46330 polyhedrin - Euxoa
718 287 2 S47479 outer surface prot
719 288 2 S32937 ureD protein - Kle
720 289 2 A42887 urease-associated
721 290 2 B69781 multidrug-efflux t
722 291 2 G71632 hypothetical prote
723 292 2 T19033 hypothetical prote
724 293 2 S62150 hypothetical prote
725 294 2 T32252 hypothetical prote
726 295 2 B72641 hypothetical prote
727 296 2 S71541 outer surface prot
728 297 2 S71541 outer surface prot
729 298 2 S71541 outer surface prot
730 299 2 I40102 outer surface prot
731 300 2 I40099 outer surface prot
732 301 2 I40097 outer surface prot
733 302 2 S23112 outer surface prot
734 303 2 S71531 outer surface prot
735 304 2 B90819 probable antitermi
736 305 2 B85627 probable antitermi
737 306 2 S83643 cytochrome-c oxida
738 307 2 B71964 enoyl-acyl carrier
739 308 2 C82717 50S ribosomal prot
740 309 2 T46825 catechol 1,2-dioxy
741 310 2 F82325 conserved hypothet
742 311 2 B86163 probable TetR-fami
743 312 2 T18742 protein F15K3.21 l
744 313 2 A70425 hypothetical prote
745 314 2 AH0460 conserved hypothet
746 315 2 AG2807 hypothetical prote
747 316 2 E97586 probable noom gene
748 317 2 AG1842 uroporphyrinogen-I
749 318 2 B38965 hypothetical prote
750 319 2 A97530 BHI246 hypothetical
751 320 2 AB2749 ABC transporter, m
752 321 2 T49784 hypothetical prote
753 322 2 I60122 rsu-1 homolog - hu
754 323 2 S5770 RSP-1 protein - mo
755 324 2 T03942 rRNA N-glycosidase
756 325 2 T11550 probable membrane
757 326 2 F86679 hypothetical prote
758 327 2 A72411 hypothetical prote

760 0.8 278 2 E72307 transposase-relate
761 0.8 279 2 C82915 ribosomal protein
762 0.8 279 2 S76797 hypothetical prote
763 0.8 279 2 G83041 probable N-hydroxy
764 0.8 280 2 T03559 2-hydroxyhepta-2,4
765 0.8 280 2 D81705 conserved hypothet
766 0.8 280 2 T25898 hypothetical prote
767 0.8 280 2 H83523 conserved hypothet
768 0.8 281 1 R5YM2C ribosomal protein
769 0.8 282 2 S16617 opacity protein op
770 0.8 282 2 D64456 protein-export mem
771 0.8 282 2 S76911 hypothetical prote
772 0.8 283 2 C84321 hypothetical prote
773 0.8 283 2 D64448 hypothetical prote
774 0.8 283 2 AC1797 Partition protein
775 0.8 283 2 AE1423 Partition protein
776 0.8 284 2 G64207 ribosomal protein
777 0.8 284 2 A75356 conserved hypothet
778 0.8 284 2 E90190 conserved hypothet
779 0.8 284 2 S18957 fix23-5 protein -
780 0.8 284 2 AD3184 pirin-like protein
781 0.8 286 2 G83378 probable short-cha
782 0.8 288 2 A86384 unknown protein [i
783 0.8 288 2 L51620 homeobox protein -
784 0.8 288 2 C72407 hypothetical prote
785 0.8 289 2 E64330 dihydrodipicolinat
786 0.8 289 2 C97239 probable membrane
787 0.8 289 2 E90397 dihydrodipicolinat
788 0.8 290 2 T23416 hypothetical prote
789 0.8 290 2 H83467 probable transcrip
790 0.8 291 2 H81024 Arp synthase Fl, 9
791 0.8 291 2 B81970 H+transporting tw
792 0.8 291 2 G83660 conserved hypothet
793 0.8 291 2 S66101 hypothetical prote
794 0.8 292 2 T06201 xyloglucan endo-1,
795 0.8 292 2 T26344 hypothetical prote
796 0.8 293 2 D95859 probable dihydrodi
797 0.8 293 2 A89818 hypothetical prote
798 0.8 293 2 D64984 hypothetical trans
799 0.8 293 2 A98010 probable transcrip
800 0.8 293 2 C85854 probable transcrip
801 0.8 294 2 B72246 dihydrodipicolinat
802 0.8 294 2 S72322 ribosomal protein
803 0.8 294 2 S05328 inner membrane pro
804 0.8 294 2 AE0780 cytidine deaminase
805 0.8 294 2 H96662 hypothetical prote
806 0.8 294 2 S19226 cold-regulated pro
807 0.8 295 2 C69180 adhesion protein -
808 0.8 295 2 T22833 hypothetical prote
809 0.8 296 2 AH3343 serine O-acetyltra
810 0.8 296 2 A83109 hypothetical prote
811 0.8 296 2 E69500 molybdopterin bios
812 0.8 296 2 T29892 hypothetical prote
813 0.8 296 2 S52254 copper resistance
814 0.8 296 2 H95404 hypothetical prote
815 0.8 296 2 H75557 hypothetical prote
816 0.8 296 2 AD2434 hypothetical prote
817 0.8 297 2 E36002 probable sugar upt
818 0.8 297 2 T13264 repressor protein
819 0.8 297 2 T36724 probable membrane
820 0.8 298 1 F69152 ribokinase - Metha
821 0.8 298 2 S53761 triose-phosphate i
822 0.8 298 2 D75481 aldose epimerase f
823 0.8 299 2 B95149 heat shock protein
824 0.8 299 2 T72210 hypothetical prote
825 0.8 299 2 T26365 hypothetical prote
826 0.8 299 2 A83421 peptidoglycan bind
827 0.8 299 2 AE3153 transcription regu
828 0.8 300 1 S40827 probable sugar kin
829 0.8 300 1 H70557 probable acyl-CoA
830 0.8 300 2 T01598 hypothetical prote
831 0.8 300 2 S69028 probable membrane
832 0.8 301 1 RLZMFI rRNA N-glycosidase

833	6	0.8	302	2	A93017	heat shock protein	906	6	0.8	321	2	C84664	epoxide hydrolase
834	6	0.8	302	2	S11860	rRNA N-glycosidase	907	6	0.8	321	2	T07043	probable epoxide h
835	6	0.8	302	2	D82996	probable transcrip	908	6	0.8	321	2	JC5460	intracellular alka
836	6	0.8	302	2	T48871	catechol 1,2-dioxy	909	6	0.8	321	2	F82296	conserved hypotet
837	6	0.8	302	2	S50609	hypothetical prote	910	6	0.8	321	2	H83560	conserved type II s
838	6	0.8	302	2	S34390	hypothetical prote	911	6	0.8	322	2	G83922	intracellular alka
839	6	0.8	302	2	A90572	aquaporin-like tra	912	6	0.8	322	2	G82018	lipopolysaccharide
840	6	0.8	302	2	D83268	conserved hypotet	913	6	0.8	322	2	B81001	heptosyltransferas
841	6	0.8	302	2	A95387	protein imported	914	6	0.8	322	2	T14597	proteinase homolog
842	6	0.8	303	1	A55592	cheA activity-modu	915	6	0.8	323	2	A55983	microtubule-associ
843	6	0.8	303	2	H97212	co/Zn/cd efflux sy	916	6	0.8	323	2	E36144	cobd protein - pse
844	6	0.8	304	2	S11859	rRNA N-glycosidase	917	6	0.8	323	2	S47966	probable lipid tra
845	6	0.8	304	2	T18005	hypothetical prote	918	6	0.8	323	2	D82987	hypothetical prote
846	6	0.8	304	2	E70698	hypothetical prote	919	6	0.8	324	2	A46631	lactose-binding le
847	6	0.8	305	2	S23711	olfactory factor O	920	6	0.8	324	2	T31992	hypothetical prote
848	6	0.8	305	2	H47099	nitrogen assimilat	921	6	0.8	324	2	T49143	hypothetical prote
849	6	0.8	305	2	H90976	nitrogen assimilat	922	6	0.8	324	2	T18763	hypothetical prote
850	6	0.8	305	2	F85823	nitrogen assimilat	923	6	0.8	324	2	A31920	collagen sgt-1 pre
851	6	0.8	305	2	D64963	nitrogen assimilat	924	6	0.8	324	2	S48122	transcription fact
852	6	0.8	306	2	T50038	beta-1.3-N-acetyl	925	6	0.8	325	2	E83387	copper resistance
853	6	0.8	306	2	T70657	probable mr - Myc	926	6	0.8	325	2	A82281	ferric vibriobacti
854	6	0.8	307	2	AH3112	hypothetical prote	927	6	0.8	325	2	H70073	two-component sens
855	6	0.8	307	2	E98174	oligopeptide ABC t	928	6	0.8	325	2	T36562	probable ion-trans
856	6	0.8	308	2	D82897	conserved hypotet	929	6	0.8	326	2	T04344	peroxidase (EC 1.1
857	6	0.8	308	2	E64534	conserved hypotet	930	6	0.8	326	2	AC0878	Type II secretion,
858	6	0.8	308	2	C71972	hypothetical prote	931	6	0.8	327	2	S53305	ferredoxin-NADP re
859	6	0.8	309	2	S76393	hypothetical prote	932	6	0.8	327	2	AC0156	probable cobalamin
860	6	0.8	309	2	D70464	hypothetical prote	933	6	0.8	327	2	T18766	hypothetical prote
861	6	0.8	309	2	AH1528	conserved hypotet	934	6	0.8	327	2	T35482	hypothetical prote
862	6	0.8	309	2	AF1171	conserved hypotet	935	6	0.8	327	2	F72482	hypothetical prote
863	6	0.8	309	2	E75565	hypothetical prote	936	6	0.8	327	2	F95889	probable dehydroge
864	6	0.8	309	2	T36297	probable DNA-bindi	937	6	0.8	328	2	AE3312	ribose-phosphate d
865	6	0.8	310	1	JU0299	ADPglyceronanno-he	938	6	0.8	328	2	A91012	hypothetical prote
866	6	0.8	310	2	E90231	UDP-glucose 4-epim	939	6	0.8	328	2	C85856	hypothetical prote
867	6	0.8	310	2	A98191	ADP-L-glycero-D-ma	940	6	0.8	328	2	D64986	yeir protein - Esc
868	6	0.8	310	2	H86038	ADP-L-glycero-D-ma	941	6	0.8	328	2	AD0784	conserved hypotet
869	6	0.8	310	2	F90011	conserved hypotet	942	6	0.8	328	2	T19815	hypothetical prote
870	6	0.8	311	1	D69219	integrase-recombin	943	6	0.8	328	2	T42996	hypothetical prote
871	6	0.8	311	2	AF2668	deacetylase [impor	944	6	0.8	329	1	A36952	hypothetical prote
872	6	0.8	311	2	D97450	histone deacetylase	945	6	0.8	329	2	AE0378	CDP-6-deoxy-delta-
873	6	0.8	311	2	F86341	hypothetical prote	946	6	0.8	329	2	F82922	phosphate transpor
874	6	0.8	312	1	G87328	ISCC2, transposase	947	6	0.8	329	2	C75400	geranylgeranyl dip
875	6	0.8	312	1	Z2BPB4	rIIB protein - pha	948	6	0.8	329	2	T44459	arginine metabolism
876	6	0.8	312	2	T35400	probable phytoene	949	6	0.8	329	2	A83405	probable hydroxyla
877	6	0.8	312	2	A64084	lysophospholipase	950	6	0.8	329	2	AD1083	dinitrogenase redu
878	6	0.8	313	2	D95866	probable transcrip	951	6	0.8	330	2	G70792	hypothetical prote
879	6	0.8	313	2	T21966	hypothetical prote	952	6	0.8	330	2	B70849	hypothetical prote
880	6	0.8	313	2	AF1973	hypothetical prote	953	6	0.8	330	2	S39588	peptide transport
881	6	0.8	314	2	T36913	probable integral	954	6	0.8	330	2	AH0656	peptide transport
882	6	0.8	314	2	D87576	oxidoreductase, al	955	6	0.8	330	2	S35439	transcription fact
883	6	0.8	314	2	D90457	cytochrome b558/56	956	6	0.8	330	2	AE2535	hypothetical prote
884	6	0.8	314	2	T28879	hypothetical prote	957	6	0.8	331	1	DECHLM	L-lactate dehydrog
885	6	0.8	315	2	AG1131	B. subtilis Yeac h	958	6	0.8	331	2	F88638	protein F58P6.6 [i
886	6	0.8	315	2	AG1491	conserved hypotet	959	6	0.8	331	2	G86671	hypothetical prote
887	6	0.8	315	2	JCS501	chemoreceptor TB56	960	6	0.8	331	2	T28208	hypothetical prote
888	6	0.8	315	2	AD3547	ribose transport s	961	6	0.8	332	2	S12151	l-lactate dehydrog
889	6	0.8	315	2	AF3060	conserved hypotet	962	6	0.8	332	2	B86176	protein F19P19.13
890	6	0.8	315	2	B98226	hypothetical prote	963	6	0.8	332	2	T12433	malate dehydrogena
891	6	0.8	315	2	E83874	nickel ABC transpo	964	6	0.8	332	2	T02935	malate dehydrogena
892	6	0.8	316	2	T34023	hypothetical prote	965	6	0.8	332	2	T09291	malate dehydrogena
893	6	0.8	317	2	H75453	electron transfer	966	6	0.8	332	2	T23903	hypothetical prote
894	6	0.8	317	2	T39869	probable lysophosp	967	6	0.8	332	2	B81383	hypothetical prote
895	6	0.8	318	2	T49714	related to spliceo	968	6	0.8	333	2	E95212	probable lipoprote
896	6	0.8	318	1	ERADF3	fiber protein - hu	969	6	0.8	333	2	E98076	hypothetical prote
897	6	0.8	319	2	I60446	Shiga-like cytotox	970	6	0.8	334	2	T36052	catabolite control
898	6	0.8	319	2	S52807	fiber protein - hu	971	6	0.8	334	2	H69148	hypothetical prote
899	6	0.8	319	2	A85437	hypothetical prote	972	6	0.8	334	2	E69361	signal-transducing
900	6	0.8	320	1	DAAL2E	catechol 2,3-dioxy	973	6	0.8	334	2	AG3029	hypothetical prote
901	6	0.8	320	2	D84664	probable epoxide h	974	6	0.8	334	2	B87545	hypothetical prote
902	6	0.8	320	2	B82520	hypothetical prote	975	6	0.8	334	2	E86794	hypothetical prote
903	6	0.8	320	2	AB3630	taurine-binding pe	976	6	0.8	334	2	G83533	hypothetical prote
904	6	0.8	321	1	S58614	cytochrome c-type	977	6	0.8	334	2	T20562	hypothetical prote
905	6	0.8	321	1	JQ0288	cytochrome c-type	978	6	0.8	334	2	B97789	hypothetical prote

979 6 0.8 335 1 XXFOGA
 980 6 0.8 335 2 S31625
 981 6 0.8 335 2 S81736
 982 6 0.8 335 2 T25498
 983 6 0.8 335 2 T34296
 984 6 0.8 336 1 S75272
 985 6 0.8 336 2 H96799
 986 6 0.8 336 2 B90071
 987 6 0.8 336 2 T44988
 988 6 0.8 336 2 B72653
 989 6 0.8 337 2 T23591
 990 6 0.8 338 2 S12731
 991 6 0.8 338 2 D84249
 992 6 0.8 338 2 B64148
 993 6 0.8 338 2 C64389
 994 6 0.8 339 2 B36868
 995 6 0.8 339 2 T16273
 996 6 0.8 340 1 OKBYR1
 997 6 0.8 340 2 S69194
 998 6 0.8 340 2 E70714
 999 6 0.8 340 2 T32646
 1000 6 0.8 340 2 T05120

phosphatidylcholin
 glycerophospholipi
 phospho-N-acetylmu
 hypothetical prote
 hypothetical prote
 cytochrome d ubiq
 ornithine transcar
 oxidoreductase [im
 hypothetical prote
 cysteine synthase
 L-serine ammonia-1
 alkanal monooxygen
 hypothetical prote
 N5,N10-methylenete
 copB homolog - Xan
 hypothetical prote
 protein kinase byr
 N4-(beta-N-acetylgl
 probable GDP-D-man
 hypothetical prote
 hypothetical prote

ALIGNMENTS

RESULT 1
 D81883
 Probable outer-membrane receptor protein NMA1161 [imported] - Neisseria meningitidis (str
 C;Species: Neisseria meningitidis
 C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C;Accession: D81883
 R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A;Reference number: A81775; MUID:20222556; PMID:10761919
 A;Accession: D81883
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-764 <PAR>
 A;Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CAB84423.1; PID:G737985
 A;Experimental source: serogroup A, strain Z2491
 C;Genetics:
 A;Gene: NMA1161

Query Match 23.1%; Score 175; DB 2; Length 764;
 Best Local Similarity 100.0%; Pred. No. 3.1e-174; Indels 0; Gaps 0;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPLGAHQRTARFALSGNNYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTEVG 532
 Db 479 FLPLGAHQRTARFALSGNNYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTEVG 538

Qy 533 NKHLNKSNNIELALGYEGDRWQYNLALYRNFRGNYIAQTLDNGRGPKEIDDSMKL 592
 Db 539 NKHLNKSNNIELALGYEGDRWQYNLALYRNFRGNYIAQTLDNGRGPKEIDDSMKL 598

Qy 593 VRYNQSGADFYGAEGEYIFKPTPRYRIGVSGDYVGRGLKNLPSLPGRDAYGNRP 647
 Db 599 VRYNQSGADFYGAEGEYIFKPTPRYRIGVSGDYVGRGLKNLPSLPGRDAYGNRP 653

RESULT 2
 A72479
 hypothetical protein APE2473 - Aeropyrum pernix (strain K1)
 C;Species: Aeropyrum pernix
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C;Accession: A72479
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerop
 A;Reference number: A72450; MUID:99310339; PMID:10382966
 A;Accession: A72479
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-205 <KAW>
 A;Cross-references: DDBJ:AP000064; NID:G5105945; PIDN:BAAS1489.1; PID:G5106178
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE2473
 C;Superfamily: Aeropyrum pernix hypothetical protein APE2473

Query Match 1.2%; Score 9; DB 2; Length 205;
 Best Local Similarity 100.0%; Pred. No. 0.63;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 318 HAHAHNGKP 326
 Db 162 HAHAHNGKP 170

RESULT 3
 T05478
 Peroxidase (EC 1.11.1.7) prxr1 - Arabidopsis thaliana
 N;Alternate names: protein T805.170
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 26-Aug-1999
 C;Accession: T05478
 R;Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuel
 submitted to the Protein Sequence Database, February 1998
 A;Reference number: Z15417
 A;Accession: T05478
 A;Molecule type: DNA
 A;Residues: 1-323 <BEV>
 A;Cross-references: EMBL:AL021890
 A;Experimental source: cultivar Columbia; BAC clone T805
 C;Genetics:
 A;Gene: prxr1
 A;Map position: 4
 A;Introns: 69/3; 132/3; 188/1
 A;Note: T805.170
 C;Superfamily: peroxidase
 C;Keywords: oxidoreductase

Query Match 1.2%; Score 9; DB 2; Length 323;
 Best Local Similarity 100.0%; Pred. No. 0.97;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 408 ALSATSEAV 416
 Db 9 ALSATSEAV 17

RESULT 4
 S54307
 myosin heavy chain - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 02-Mar-2001
 C;Accession: S54307
 R;Reinhard, J.; Scheel, A.A.; Diekmann, D.; Hall, A.; Ruppert, C.; Baehler, M.
 EMBO J. 14, 697-704, 1995
 A;Title: A novel type of myosin implicated in signalling by rho family GTPases.
 A;Reference number: S54307; MUID:95188874; PMID:7882973
 A;Accession: S54307
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-1980 <REI>
 A;Cross-references: EMBL:X77609; NID:G639998; PIDN:CAAS4700.1; PID:G639999
 C;Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase C
 C;Keywords: nucleotide binding; P-loop
 F;149-942/Domain: myosin motor domain homology <MMOT>
 F;239-246/Region: nucleotide-binding motif A (P-loop)
 F;1593-1641/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 1.2%; Score 9; DB 2; Length 1980;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSG 213
 |||||
 Db 1690 TEGLYRKSG 1698

RESULT 5
 A59256
 myosin-IXb [similarity] - human
 C:Species: Homo sapiens (man)
 C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Mar-2001
 C:Accession: A59256; I61700
 R:Wirth, J.A.; Jensen, K.A.; Post, P.L.; Bement, W.M.; Mooseker, M.S.
 J. Cell Sci. 109, 653-661, 1996
 A:Title: Human myosin-IXb, an unconventional myosin with a chimerin-like rho/rac GTPase-
 A:Reference number: A59256; MUID:97063843; PMID:8907710
 A:Accession: A59256
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-2022 <WIR>
 A:Cross-references: GB:U42391; NID:g1147782; PIDN:AAC50402.1; PID:g1147783
 R:Bement, W.M.; Hanson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
 Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
 A:Title: Identification and overlapping expression of multiple unconventional myosin ge
 A:Reference number: A55758; MUID:94294418; PMID:8022818
 A:Accession: I61700
 A:Status: preliminary; translated from GB/EMBL/DDBV
 A:Molecule type: mRNA
 A:Residues: 234-322 <RES>
 A:Cross-references: GB:I29149; NID:g457257; PIDN:AAA20912.1; PID:g531142
 C:Genetics:
 A:Gene: GDB:MYO9B; OMIM:602129
 A:Map position: 19p13.1
 C:Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase C zi
 C:Keywords: nucleotide binding; P-loop
 F:149-941/Domain: myosin motor domain homology #status atypical <WMO>
 F:239-246/Region: nucleotide-binding motif A (P-loop)

Query Match 1.2%; Score 9; DB 2; Length 2022;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSG 213
 |||||
 Db 1730 TEGLYRKSG 1738

RESULT 6
 AB1484
 hypothetical protein lin0409 [imported] - Listeria innocua (strain Clip11262)
 C:Species: Listeria innocua
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AB1484
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AB1484
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC95642.1; PID:g16412838; GSPDB:GN00178
 A:Experimental source: strain Clip11262
 C:Genetics:
 A:Gene: lin0409

Query Match 1.1%; Score 8; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTTLKPI 9
 |||||
 Db 46 AQTTLKPI 53

RESULT 7
 AH1123
 hypothetical protein lmo0391 [imported] - Listeria monocytogenes (strain EGD-e)
 C:Species: Listeria monocytogenes
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AH1123
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AH1123
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC98470.1; PID:g16409769; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0391

Query Match 1.1%; Score 8; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTTLKPI 9
 |||||
 Db 46 AQTTLKPI 53

RESULT 8
 AE2883
 nitroreductase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AE2883
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCle
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AE2883
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-196 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL43483.1; PID:g17740989; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu2496
 A:Map position: circular chromosome

Query Match 1.1%; Score 8; DB 2; Length 196;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 LSSGNLEK 187
 |||||
 Db 70 LSSGNLEK 77

```

Bioorg. Khim. 16, 324-335, 1990
A;Title: NAD-dependent formate dehydrogenase from methylotrophic bacterium Pseudomonas
A;Reference number: JU0334; MUID:90290536; PMID:2357236
A;Accession: JU0334
A;Molecule type: protein
A;Residues: 1-393 <POP>
A;Note: article in Russian with English abstract
C;Superfamily: Neurospora formate dehydrogenase
C;Keywords: homodimer; NAD; oxidoreductase

Query Match 1.1%; Score 8; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 12
T19180
hypothetical protein C10C5.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19180
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19085
A;Accession: T19180
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-399 <WIL>
A;Cross-references: EMBL:Z68214; PIDN:CAA92445.1; GSPDB:GN00022; CESP:C10C5.3
A;Experimental source: clone C10C5
C;Genetics:
A;Gene: CESP:C10C5.3
A;Map position: 4
A;Introns: 27/1; 48/3; 215/3; 268/1; 303/3; 330/2; 350/3; 383/1

Query Match 1.1%; Score 8; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 LINTPLLA 22
Db 361 LINTPLLA 368

RESULT 13
JC7815
formate dehydrogenase (EC 1.2.1.2) - Paracoccus sp. (Strain 12-A)
C;Species: Paracoccus sp.
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002
C;Accession: JC7815
R;Shinoda, T.; Satoh, T.; Mineki, S.; Iida, M.; Taguchi, H.
Biosci. Biotechnol. Biochem. 66, 271-276, 2002
A;Title: Cloning, nucleotide sequencing, and expression in Escherichia coli of the gen
A;Reference number: JC7815; PMID:1199398; MUID:21994041
A;Accession: JC7815
A;Molecule type: DNA
A;Residues: 1-400 <SHI>
A;Cross-references: DDBJ:AB071373
C;Comment: This enzyme, which catalyzes the conversion of formate to carbon dioxide wi
ti-enzyme systems, such as bioreactors involving NADH as a coenzyme, and plays a key r
C;Genetics:
A;Gene: fdh
C;Keywords: oxidoreductase

Query Match 1.1%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179

```

```

probable enzyme (AE005300) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: C97659
R;Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: C97659
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88228.1; PID:gl5157682; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4536
A;Map position: circular chromosome

Query Match 1.1%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 LSSGNLEK 187
Db 70 LSSGNLEK 77

RESULT 10
FNBSLC
beta-lactamase (EC 3.5.2.6) III precursor - Bacillus cereus
N;Alternate names: penicillinase III
C;Species: Bacillus cereus
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 12-Apr-1996
C;Accession: A27755
R;Hussain, M.; Pastor, F.I.J.; Lampen, J.O.
J. Bacteriol. 169, 579-586, 1987
A;Title: Cloning and sequencing of the blaZ gene encoding beta-lactamase III, a lipoprot
A;Reference number: A27755; MUID:87109042; PMID:3027036
A;Accession: A27755
A;Molecule type: DNA
A;Residues: 1-316 <HSN>
A;Experimental source: strain 569/H
C;Comment: This membrane-bound enzyme resembles several gram-positive class A beta-lacta
C;Comment: This enzyme is an acyl-glyceride thioether-linked lipoprotein with the hydrog
C;Genetics:
A;Gene: blaZ
C;Superfamily: beta-lactamase I
C;Keywords: antibiotic resistance; hydrolase; lipoprotein; membrane protein; penicillin
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-316/Product: beta-lactamase III #status predicted <MAT>
F;95/Active site: Ser #status predicted

Query Match 1.1%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 KPVLVLSIL 14
Db 278 KPVLVLSIL 285

RESULT 11
JU0334
formate dehydrogenase (EC 1.2.1.2) - Pseudomonas sp.
N;Alternate names: NAD-dependent formate dehydrogenase
C;Species: Pseudomonas sp.
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Sep-2002
C;Accession: JU0334
R;Popov, V.O.; Shumilin, I.A.; Ustinnikova, T.B.; Lamzin, V.S.; Egorov, T.A.

```

```

Db          54 VSGELGLR 61
|||||
RESULT 14
C95293
Probable NAD-dependent formate dehydrogenase [imported] - Sinorhizobium meliloti (strain
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 23-Sep-2002
C:Accession: C95293
R:Barnett, M.J.; Keating, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: C95293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <KUR>
A:Cross-references: GB:AE006469; PIDN:AKG4909.1; PID:gl14523329; GSPDB:GNO0165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma0478
A:Genome: plasmid
C:Superfamily: Neurospora formate dehydrogenase

Query Match          1.1%; Score 8; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          172 VSGELGLR 179
|||||
Db          56 VSGELGLR 63

RESULT 15
S76859
hypothetical protein sl11550 - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76859
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76859
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-544 <KAN>
A:Cross-references: EMBL:D90917; GB:AB001339; NID:gl653836; PIDN:BA18771.1; PID:gl65386
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG
C:Superfamily: Synecocystis hypothetical protein slr0042

Query Match          1.1%; Score 8; DB 2; Length 544;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          179 RLSSGNLE 186
|||||
Db          222 RLSSGNLE 229

RESULT 16
T40040
GTPase-activator protein for Rho-like GTPases - fission yeast (Schizosaccharomyces pom)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40040
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21901
A:Accession: T40040
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-892 <WOO>
A:Cross-references: EMBL:AL031517; PIDN:CAA20650.1; GSPDB:GNO0067; SPDB:SPBPC38E12.03
A:Experimental source: strain 972h-; cosmid c28E12
C:Genetics:
A:Gene: SPDB:SPBPC28E12.03
A:Map position: 2
A:Introns: 7/2; 124/1

Query Match          1.1%; Score 8; DB 2; Length 892;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          206 EGYLRKSG 213
|||||
Db          738 EGYLRKSG 745

RESULT 17
S61174
hypothetical protein YDR379w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D9481.4
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C:Accession: S61174
R:Ding, H.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of S. cerevisiae cosmid 9481.
A:Reference number: S61159
A:Accession: S61174
A:Molecule type: DNA
A:Residues: 1-1009 <DIN>
A:Cross-references: EMBL:U28373; NID:g849184; PIDN:AAB64815.1; PID:g849200; MIPS:YDR37;
A:Experimental source: strain S288C (AB972)
C:Genetics:
A:Gene: SGD:RGA2
A:Cross-references: SGD:S0002787; MIPS:YDR379w
A:Map position: 4R
C:Superfamily: LIM metal-binding repeat homology
F:13-66/Domain: LIM metal-binding repeat homology <LIM2>

Query Match          1.1%; Score 8; DB 2; Length 1009;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          206 EGYLRKSG 213
|||||
Db          822 EGYLRKSG 829

RESULT 18
S34567
gene L protein - hemorrhagic septicemia virus (fragment)
C:Species: hemorrhagic septicemia virus
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
C:Accession: S34567
R:Schuetze, H.
submitted to the EMBL Data Library, June 1993
A:Reference number: S34562

```

A;Accession: S34567
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-54 <SCH>
 A;Cross-references: EMBL:X3873
 C;Genetics:
 A;Gene: 1

Query Match 0.9%; Score 7; DB 2; Length 54;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 HQERLPS 513
 |||||
 Db 11 HQERLPS 17

RESULT 19

A81783
 hypothetical protein NMA2114 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
 C;Species: Neisseria meningitidis
 C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C;Accession: A81783
 R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A;Reference number: A81775; MUID:20222556; PMID:10761919
 A;Accession: A81783
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-76 <PAR>
 A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85328.1; PID:g738073
 A;Experimental source: serogroup A, strain Z2491
 C;Genetics:
 A;Gene: NMA2114

Query Match 0.9%; Score 7; DB 2; Length 76;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 DRIDANL 680
 |||||
 Db 17 DRIDANL 23

RESULT 20

E69460
 conserved hypothetical protein AF1686 - Archaeoglobus fulgidus
 C;Species: Archaeoglobus fulgidus
 C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
 C;Accession: E69460
 R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A;Reference number: A69250; MUID:98049343; PMID:9389475
 A;Accession: E69460
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-100 <KLE>
 A;Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89563.1; PID:g264886
 C;Superfamily: conserved hypothetical protein MJ0128

Query Match 0.9%; Score 7; DB 2; Length 100;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 NVAGLVD 157
 |||||

Db 57 NVAGLVD 63

RESULT 21

A72224
 hypothetical protein TM1659 - Thermotoga maritima (strain MSB8)
 C;Species: Thermotoga maritima
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C;Accession: A72224
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hick
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D
 C.M.

Nature 399, 323-329, 1999
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome s
 A;Reference number: A72200; MUID:99287316; PMID:10360571
 A;Accession: A72224
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-108 <ARN>
 A;Cross-references: GB:AE001808; GB:AE000512; NID:g4982233; PIDN:AAD36726.1; PID:g4982
 A;Experimental source: strain MSB8
 C;Genetics:
 A;Gene: TM1659
 C;Superfamily: Thermotoga maritima hypothetical protein TM1659

Query Match 0.9%; Score 7; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 VDYDNGP 306
 |||||
 Db 50 VDYDNGP 56

RESULT 22

S01185
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 3 - fruit fly (Drosophila melanoga
 C;Species: mitochondrion Drosophila melanogaster
 C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 03-Jun-2002
 C;Accession: S01185
 R;Garesse, R.
 Genetics 118, 649-663, 1988
 A;Title: Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary
 A;Reference number: S01185; MUID:88212147; PMID:3130291
 A;Accession: S01185
 A;Molecule type: DNA
 A;Residues: 1-117 <GAR>
 A;Cross-references: GB:M37275; EMBL:Y00610; NID:g336819; PIDN:AAA69709.1; PID:g894077
 A;Note: the author translated the initiation codon ATT for residue 1 as Ile
 C;Genetics:
 A;Gene: ND-3
 A;Cross-references: FlyBase:FBgn0013681
 A;Genome: mitochondrion
 A;Genetic code: SGC
 A;Start codon: ATT
 C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3
 C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 0.9%; Score 7; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 461 KALIDRE 467
 |||||
 Db 28 KALIDRE 34

RESULT 23

G25797
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 3 - fruit fly (Drosophila yakuba)
 N;Alternate names: NADH-ubiquinone oxidoreductase chain 3
 C;Species: mitochondrion Drosophila yakuba
 C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 03-Jun-2002

Fri Nov 21 10:34:25 2003

C;Accession: G25797
 R;Clary, D.O.; Wolstenholme, D.R.
 J. Mol. Evol. 22, 252-271, 1985
 A;Title: The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene
 A;Reference number: A92962; MUID:86089137; PMID:3001325
 A;Accession: G25797
 A;Molecule type: DNA
 A;Residues: 1-117 <CLA>
 A;Cross-references: GB:X03240; GB:J01400; GB:J01402; GB:J01403; GB:J01406; GB:J01408; GB:J01409
 C;Genetics:
 A;Gene: FlyBase:Dyak/mt:ND3
 A;Cross-references: FlyBase:FBgn0013204
 A;Genome: mitochondrion
 A;Genetic code: SGC4
 A;Start codon: ATT
 C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3
 C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 0.9%; Score 7; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 47; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

QY 461 KALIDRE 467
 |||||
 DB 28 KALIDRE 34

RESULT 24
 C87588
 hypothetical protein CC2738 [imported] - *Caulobacter crescentus*
 C;Species: *Caulobacter crescentus*
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 A;Accession: C87588
 R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A;Reference number: A87249; MUID:21173698; PMID:11259647
 A;Accession: C87588
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-117 <S70>
 A;Cross-references: GB:AE005673; NID:gl3424331; PIDN:AAK24703.1; GSPDB:GN00148
 C;Genetics:
 A;Gene: CC2738

Query Match 0.9%; Score 7; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 47; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

QY 408 ALSATSE 414
 |||||
 DB 31 ALSATSE 37

RESULT 25
 T05445
 hypothetical protein F7K2.90 - *Arabidopsis thaliana*
 C;Species: *Arabidopsis thaliana* (mouse-ear cross)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
 C;Accession: T05445
 R;Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Meves, H.W.; Mayer, K.F.X.; Schueller
 submitted to the Protein Sequence Database, November 1998
 A;Reference number: Z15416
 A;Accession: T05445
 A;Molecule type: DNA
 A;Residues: 1-125 <BEV>
 A;Cross-references: EMBL:AL033545
 A;Experimental source: cultivar Columbia; BAC clone F7K2
 C;Genetics:
 A;Map position: 4
 A;Introns: 73/3

A;Note: F7K2.90

Query Match 0.9%; Score 7; DB 2; Length 125;
 Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

QY 501 LSLTASH 507

|||||
 DB 8 LSLTASH 14

RESULT 26

T27080

hypothetical protein Y51A2D.10 - *Caenorhabditis elegans*C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C;Accession: T27080

R;McMurray, A.

submitted to the EMBL Data Library, January 1998

A;Reference number: Z20307

A;Accession: T27080

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-141 <WIL>

A;Cross-references: EMBL:AL021497; PIDN:CAA16408.1; GSPDB:GN00023; CESP:Y51A2D.10

A;Experimental source: clone Y51A2D

C;Genetics:

A;Gene: CESP:Y51A2D.10

A;Map position: 5

A;Introns: 54/1; 96/3

C;Superfamily: *Caenorhabditis* hypothetical protein C40H1.5

Query Match 0.9%; Score 7; DB 2; Length 141;

Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 TTVVVGK 41

|||||
 DB 27 TTVVVGK 33

RESULT 27

T27071

hypothetical protein Y51A2D.9 - *Caenorhabditis elegans*C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C;Accession: T27071

R;McMurray, A.

submitted to the EMBL Data Library, January 1998

A;Reference number: Z20307

A;Accession: T27071

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-141 <WIL>

A;Cross-references: EMBL:AL021497; PIDN:CAA16399.1; GSPDB:GN00023; CESP:Y51A2D.9

A;Experimental source: clone Y51A2D

C;Genetics:

A;Gene: CESP:Y51A2D.9

A;Map position: 5

A;Introns: 54/1; 96/3

C;Superfamily: *Caenorhabditis* hypothetical protein C40H1.5

Query Match 0.9%; Score 7; DB 2; Length 141;

Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 TTVVVGK 41

|||||
 DB 27 TTVVVGK 33

RESULT 28

T35846

probable integral membrane protein - Streptomyces coelicolor (fragment)
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 03-Nov-2000
 C/Accession: T35846
 R/Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M. submitted to the EMBL Data Library, April 1999
 A/Reference number: Z21591
 A/Accession: T35846
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-149 <SAU>
 A/Cross-references: EMBL:AL049727; PIDN:CAB41547.1; GSPDB:GN000070; SCODEB:SC9B1.01c
 A/Experimental source: strain A3(2)
 C/Genetics:
 A/Gene: SCODEB:SC9B1.01c
 C/Superfamily: integral membrane protein HP0228

Query Match 0.9%; Score 7; DB 2; Length 149;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 DOYGLPA 264
 |||||
 Db 83 DOYGLPA 89

RESULT 29
 DTECR
 aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Escherichia coli (strain N/Alternate names: aspartate transcarbamylase; carbamylaspartotranskinase
 C/Species: Escherichia coli
 C/Date: 24-Apr-1984 #sequence_revision 03-Aug-1984 #text_change 01-Mar-2002
 C/Accession: A93985; A00560; S56470; S56404; G65236
 R/Schachman, H.K.; Pauza, C.D.; Navre, M.; Karels, M.J.; Wu, L.; Yang, Y.R. Proc. Natl. Acad. Sci. U.S.A. 81, 1115-1119, 1984
 A/Title: Location of amino acid alterations in mutants of aspartate transcarbamoylase: S
 A/Reference number: A93985; MUID:84119419; PMID:6364131
 A/Accession: A93985
 A/Molecule type: DNA
 A/Residues: 1-153 <SCH>
 A/Cross-references: GB:K01472; NID:gl47463; PIDN:AAA24477.1; PID:gl47465
 R/Weber, K. Nature 218, 1116-1119, 1968
 A/Title: New structural model of Escherichia coli aspartate transcarbamylase and the ami
 A/Reference number: A93154; MUID:68284659; PMID:4872216
 A/Accession: A00560
 A/Molecule type: Protein
 A/Residues: 1-3; ND', 6-9, 'AE', 12-18, 'N', 20-23, 'E', 25-38, 'QD', 41-86, 'ND', 89-102, 'NID', 106
 R/Monaco, H.L.; Crawford, J.L.; Lipscomb, W.N. Proc. Natl. Acad. Sci. U.S.A. 75, 5276-5280, 1978
 A/Title: Three-dimensional structures of aspartate carbamoyltransferase from Escherichia
 A/Reference number: A93823; MUID:79074799; PMID:364472
 C/Contents: annotation; X-ray crystallography, 2.8 angstroms
 R/Xe, H.; Honzatko, R.B.; Lipscomb, W.N. Proc. Natl. Acad. Sci. U.S.A. 81, 4037-4040, 1984
 A/Title: Structure of unligated aspartate carbamoyltransferase of Escherichia coli at 2.
 A/Reference number: A93993; MUID:84248054; PMID:6377306
 C/Contents: annotation; X-ray crystallography, 2.6 angstroms; quaternary structure
 R/Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R. Nucleic Acids Res. 23, 2105-2119, 1995
 A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
 A/Reference number: S56314; MUID:95334362; PMID:7610040
 A/Accession: S56470
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-153 <BUR>
 A/Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97141.1; PID:g537086
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R/Cunin, R.; Jacobs, A.; Charlier, D.; Crabeel, M.; Herve, G.; Glansdorff, N.; Pierard, J. Mol. Biol. 186, 707-713, 1985
 A/Title: Structure-function relationship in allosteric aspartate carbamoyltransferase fr
 A/Reference number: I56404; MUID:86143826; PMID:3912513
 A/Accession: I56404

A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 128-153 <RES>
 A/Cross-references: GB:M28578; NID:g147480; PIDN:AAA24487.1; PID:g147481
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: G65236
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-153 <BLAT>
 A/Cross-references: GB:AE000495; GB:U00096; NID:g2367361; PIDN:AACT7201.1; PID:g179069
 A/Experimental source: strain K-12, substrain MG1655
 C/Genetics:
 A/Gene: pyrI
 A/Map position: 97 min
 C/Complex: heterododecamer of two trimers of catalytic chains and three dimers of regu
 C/Function:
 A/Description: catalyzes the transcarbamylation of carbanoyl phosphate and aspartate, t
 A/Pathway: pyrimidine nucleotide biosynthesis
 C/Superfamily: aspartate carbamoyltransferase regulatory chain
 C/Keywords: acyltransferase; heterododecamer; homodimer; homohehexamer; homotrimer; pyrI
 F109,114,138,141/Binding site: zinc (Cys) #status experimental

Query Match 0.9%; Score 7; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
 |||||
 Db 91 VVGKSRP 97

RESULT 30
 DTEBCT
 aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Salmonella typhimurium
 C/Species: Salmonella typhimurium
 C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
 C/Accession: S00050
 R/Michaels, G.; Keiln, R.A.; Nargang, F.E. Eur. J. Biochem. 166, 55-61, 1987
 A/Title: Cloning, nucleotide sequence and expression of the pyrBI operon of Salmonella
 A/Reference number: S00028; MUID:87246692; PMID:3036524
 A/Accession: S00050
 A/Molecule type: DNA
 A/Residues: 1-153 <MIC>
 A/Cross-references: GB:X05641; NID:g47861; PIDN:CAA29130.1; PID:g47864
 C/Comment: the active enzyme contains two trimers of catalytic chains and three dimers
 C/Genetics:
 A/Gene: pyrI
 C/Superfamily: aspartate carbamoyltransferase regulatory chain
 C/Keywords: heterododecamer; homodimer; homohehexamer; pyrimidine nucleotide biosynthesi
 F109,114,138,141/Binding site: zinc (Cys) #status predicted

Query Match 0.9%; Score 7; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
 |||||
 Db 91 VVGKSRP 97

RESULT 31
 E91281
 aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Escherichia coli (strai
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 31-Dec-2001
 C/Accession: E91281
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehi, K.; Yokoyama, K.; Han, C. Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E91281
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA038644.1; PID:gl3364698; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5221
C;Superfamily: aspartate carbamoyltransferase regulatory chain
C;Keywords: transferase

Query Match 0.9%; Score 7; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
Db 91 VVGKSRP 97
|||||

RESULT 32
F86122
aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Dec-2001
C;Accession: F86122
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206951
A;Accession: F86122
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <SNT>
A;Cross-references: GB:AE005174; NID:gl2519252; PIDN:AAG59442.1; GSPDB:GN00145; UWGP:Z58
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: pyrI
C;Superfamily: aspartate carbamoyltransferase regulatory chain
C;Keywords: transferase

Query Match 0.9%; Score 7; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
Db 91 VVGKSRP 97
|||||

RESULT 33
AC1059
aspartate carbamoyltransferase regulatory chain pyrI [imported] - Salmonella enterica su
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC1059
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC1059
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD06920.1; PID:g16505568; GSPDB:GN00176

C;Genetics:
A;Gene: pyrI
C;Superfamily: aspartate carbamoyltransferase regulatory chain

Query Match 0.9%; Score 7; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
Db 91 VVGKSRP 97
|||||

RESULT 34
S76912
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C;Species: Synecocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S76912
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocyst
S.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76912
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155 <KAN>
A;Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAA18824.1; PID:d1019
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

Query Match 0.9%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 SGNLEKL 188
Db 146 SGNLEKL 152
|||||

RESULT 35
T07089
dehydrin - soybean
C;Species: Glycine max (soybean)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Aug-1999
C;Accession: T07089
R;Hsu, T.F.; Tsai, F.Y.; Hsing, Y.I.; Chow, T.Y.
submitted to the EMBL Data Library, May 1997
A;Description: Glycine max mRNA for dehydrin.
A;Reference number: Z15912
A;Accession: T07089
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-166 <HSU>
A;Cross-references: EMBL:AF004807; NID:g2270989; PIDN:AAB71225.1; PID:g2270990
A;Experimental source: strain Shi-shi; cotyledon
C;Genetics:
A;Gene: PM12
C;Superfamily: dehydrin-like protein

Query Match 0.9%; Score 7; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 YGGGASA 94
Db 59 YGGGASA 65
|||||

```

RESULT 36
A:Title: acetyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
A:Reference number: AF2864
A:Accession: AF2864
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AF2864
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF2864
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <KUR>
A:CROSS-references: GB:AE008688; PIDN:RAL4332.1; PID:gl7740825; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2343
A:Map position: circular chromosome

Query Match 0.9%; Score 7; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 LTEDVD 301
|||||
Db 9 LTEDVD 15

RESULT 37
A:Title: hypothetical protein AGR_C_4257 [imported] - Agrobacterium tumefaciens (strain C58, Cere
E97641
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: E97641
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E97641
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <KUR>
A:CROSS-references: GB:AE007869; PIDN:AAK88086.1; PID:gl15157514; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4257
A:Map position: circular chromosome

Query Match 0.9%; Score 7; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 LTEDVD 301
|||||
Db 9 LTEDVD 15

RESULT 38
C70228
conserved hypothetical protein BBF03 - Lyme disease spirochete plasmid F/1p28-1
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: C70228
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.

```

```

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: C70228
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <KLE>
A:CROSS-references: GB:AE000794; NID:g2689981; PIDN:AAK66368.1; PID:g2689985; TIGR:BBF
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 0.9%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PIVLSIL 14
|||||
Db 154 PIVLSIL 160

RESULT 39
A87609
hypothetical protein CC2907 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87609
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87609
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <STO>
A:CROSS-references: GB:AE005673; NID:gl33424527; PIDN:AAK24869.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2907

Query Match 0.9%; Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 GDALGV 80
|||||
Db 122 GDALGV 128

RESULT 40
A41740
H+-transporting two-sector ATPase (BC 3.6.3.14) delta' chain precursor - sweet potato
C:Species: Ipomoea batatas (sweet potato)
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 03-Jun-2002
C:Accession: A41740; C33306
R:MoriKani, A.; Aiso, K.; Asahi, T.; Nakamura, K.
J. Biol. Chem. 267, 72-76, 1992
A:Title: The delta'-subunit of higher plant six-subunit mitochondrial F-1-ATPase is ho
A:Reference number: A41740; MUID:92112727; PMID:1370454
A:Accession: A41740
A:Molecule type: mRNA
A:Residues: 1-200 <MOR>
A:CROSS-references: GB:D10660; GB:D90473; NID:g217937; PIDN:BA01511.1; PID:g217938
R:Kimura, T.; Nakamura, K.; Kajiuira, H.; Hattori, H.; Nelson, N.; Asahi, T.
J. Biol. Chem. 264, 3183-3186, 1989
A:Title: Correspondence of minor subunits of plant mitochondrial F-1-ATPase to F-1F-0AT
A:Reference number: A33306; MUID:89123436; PMID:2536736
A:Accession: C33306
A:Molecule type: protein
A:Residues: 'X',23-38,'X',40-50,'X',52,'X',54,'X',56 <KIT>
A:Superfamily: H+-transporting ATP synthase epsilon chain
C:Keywords: hydrolase
F-1-21/Domain: transit peptide (mitochondrion) #status predicted <TRP>

```

F:22-200/Product: H+-transporting ATP synthase delta' chain #status predicted <MAT>

Query Match 0.9%; Score 7; DB 2; Length 200;
Best Local Similarity 100.0%; Pred.No.77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 DRIDANL 680
Db 154 DRIDANL 160

Search completed: November 14, 2003, 11:08:47
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:04:59 ; Search time 42 Seconds
(without alignments)
4657.235 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTTLKPIVLSILLINTPL.....FLSDTPQMGRSFTGGVNVKF 758

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	286	37.7	758	16 Q9JZN9	Q9jzn9 neisseria m
2	175	23.1	764	16 Q9JUS2	Q9jue2 neisseria m
3	12	1.6	809	16 Q9CJW8	Q9clw8 pasteurella
4	9	1.2	205	17 Q9Y911	Q9y911 aeropyrum p
5	9	1.2	766	2 Q9AER9	Q9aer9 pasteurella
6	8	1.1	93	2 Q9ZH42	Q9zh42 morganella
7	8	1.1	99	16 Q92EP9	Q92ep9 listeria in
8	8	1.1	99	16 Q8Y9X6	Q8y9x6 listeria mo
9	8	1.1	171	11 Q9DAA9	Q9daa9 mus musculus
10	8	1.1	268	5 Q9VNN6	Q9vnn6 drosophila
11	8	1.1	268	5 Q9NH68	Q9nh68 drosophila
12	8	1.1	398	2 Q9F7P9	Q9f7p9 uncultured
13	8	1.1	399	2 Q93UW1	Q93uw1 hyphomicrob
14	8	1.1	399	5 Q17898	Q17898 caenorhabdi
15	8	1.1	400	2 Q9R5V0	Q9r5v0 mycobacteri
16	8	1.1	400	2 Q93GW3	Q93gw3 paracoccus

17	8	1.1	401	2 Q93GV1	Q93gv1 mycobacteri
18	8	1.1	401	16 Q930E7	Q930e7 rhizobium m
19	8	1.1	402	2 Q08375	Q08375 moraxella s
20	8	1.1	427	2 Q9F771	Q9f771 pseudomonas
21	8	1.1	544	16 P74655	P74655 synechocyst
22	8	1.1	925	16 Q9CWR7	Q9cmr7 pasteurella
23	8	1.1	933	3 Q74360	Q74360 schizosacch
24	8	1.1	984	16 Q8PD15	Q8pd15 xanthomonas
25	8	1.1	1009	3 Q06407	Q06407 saccharomyc
26	8	1.1	2319	10 Q8H614	Q8h614 zea mays (m
27	8	1.1	5017	2 Q8GBX6	Q8gbx6 polyangium
28	7	0.9	48	2 Q8KY74	Q8ky74 mycobacteri
29	7	0.9	48	2 Q8KY83	Q8ky83 mycobacteri
30	7	0.9	48	2 Q8KHQ1	Q8khd1 mycobacteri
31	7	0.9	55	12 Q08482	Q08482 viral hemor
32	7	0.9	60	10 Q8L411	Q8l411 oryza sativ
33	7	0.9	76	16 Q9JSV8	Q9jsv8 neisseria m
34	7	0.9	86	2 Q8KY84	Q8ky84 mycobacteri
35	7	0.9	86	2 Q8KY71	Q8ky71 mycobacteri
36	7	0.9	86	2 Q8KY78	Q8ky78 mycobacteri
37	7	0.9	86	2 Q8KY80	Q8ky80 mycobacteri
38	7	0.9	86	2 Q8KY79	Q8ky79 mycobacteri
39	7	0.9	86	2 Q8KY77	Q8ky77 mycobacteri
40	7	0.9	91	5 Q9BK93	Q9bk93 cryptospori
41	7	0.9	91	5 Q9BK94	Q9bk94 cryptospori
42	7	0.9	94	16 Q8PMD7	Q8pmd7 xanthomonas
43	7	0.9	100	17 Q28587	Q28587 archaeoglob
44	7	0.9	105	2 Q9RZW7	Q9rzw7 borrelia bu
45	7	0.9	108	16 Q9X1Y9	Q9x1y9 thermotoga
46	7	0.9	117	8 Q9MD67	Q9md67 drosophila
47	7	0.9	117	8 Q9MGW8	Q9mgw8 drosophila
48	7	0.9	117	8 Q9MGL9	Q9mgl9 drosophila
49	7	0.9	117	16 Q9A4T9	Q9a4t9 caulobacter
50	7	0.9	124	12 Q9J521	Q9j521 fowlpox vir
51	7	0.9	125	10 Q9SUW8	Q9suw8 arabidopsis
52	7	0.9	138	13 Q9DEZ7	Q9dez7 scyllorhinu
53	7	0.9	141	5 Q9XXR4	Q9xxr4 caenorhabdi
54	7	0.9	141	5 Q9XXQ6	Q9xxq6 caenorhabdi
55	7	0.9	141	10 Q8H5V0	Q8h5v0 oryza sativ
56	7	0.9	144	5 Q61831	Q61831 caenorhabdi
57	7	0.9	149	11 Q8C542	Q8c542 mus musculu
58	7	0.9	155	16 P74705	P74705 synechocyst
59	7	0.9	159	16 Q9KY20	Q9ky20 streptomyce
60	7	0.9	162	6 Q8SP19	Q8spt9 macaca mula
61	7	0.9	163	5 Q8MXY4	Q8mxy4 cryptospori
62	7	0.9	166	10 Q23957	Q23957 glycine max
63	7	0.9	169	5 Q9NC83	Q9nc83 strongyloce
64	7	0.9	171	16 Q8UCY5	Q8uecy5 agrobacteri
65	7	0.9	173	16 Q51013	Q51013 borrelia bu
66	7	0.9	184	16 Q8DD61	Q8dd61 vibrio vuln
67	7	0.9	185	16 Q9A4C9	Q9a4c9 caulobacter
68	7	0.9	189	11 Q98G88	Q98g88 rhizobium l
69	7	0.9	198	11 Q8CAH4	Q8cah4 mus musculu
70	7	0.9	201	16 Q9X0Q3	Q9x0q3 thermotoga
71	7	0.9	202	5 Q9U8Y6	Q9u8y6 tachypleus
72	7	0.9	202	5 Q9U8Y8	Q9u8y8 tachypleus
73	7	0.9	202	5 Q9U8Y7	Q9u8y7 tachypleus
74	7	0.9	202	17 Q9Y8Y6	Q9yey6 aeropyrum p
75	7	0.9	207	15 Q71842	Q71842 human immun
76	7	0.9	207	15 Q71841	Q71841 human immun
77	7	0.9	215	5 Q9TZD2	Q9trzd2 caenorhabdi
78	7	0.9	222	5 Q16539	Q16539 caenorhabdi
79	7	0.9	232	2 Q9R6S5	Q9r6s5 synechococc
80	7	0.9	232	2 Q8GAA2	Q8ga2 synechococc
81	7	0.9	232	4 Q8IX63	Q8ix63 homo sapien
82	7	0.9	234	16 Q8XIF6	Q8xif6 clostridium
83	7	0.9	234	17 Q97WB6	Q97wb6 sulfolobus
84	7	0.9	234	17 Q96XJ2	Q96xj2 sulfolobus
85	7	0.9	237	10 Q8RZY8	Q8rzy8 oryza sativ
86	7	0.9	241	16 Q99U19	Q99ul9 staphylococ
87	7	0.9	241	16 Q8CSH9	Q8csh9 staphylococ
88	7	0.9	243	4 Q96CG8	Q96cgb homo sapien
89	7	0.9	243	16 Q8XMM3	Q8xmm3 clostridium

90	7	0.9	244	16	Q9A100	Q9A100 streptococc	163	7	0.9	357	16	Q8DCE3	Q8dce3 vibrio vuln
91	7	0.9	244	16	Q31614	Q31614 bacillus su	164	7	0.9	360	5	Q8MRE6	Q8mre6 drosophila
92	7	0.9	244	16	Q8P215	Q8p215 streptococc	165	7	0.9	369	16	Q8DBH5	Q8dbh5 vibrio vuln
93	7	0.9	244	16	Q8K8B0	Q8k8b0 streptococc	166	7	0.9	375	4	Q8EGD2	Q8egd2 homo sapien
94	7	0.9	245	11	Q9D1D6	Q9d1d6 mus musculu	167	7	0.9	375	17	Q96Y17	Q96y17 sulfolobus
95	7	0.9	245	11	Q8CG08	Q8cg08 rattus norv	168	7	0.9	380	16	Q8DDZ5	Q8ddz5 vibrio vuln
96	7	0.9	245	16	Q9A604	Q9a604 caulobacter	169	7	0.9	391	16	Q8YKV3	Q8ykv3 anabaena sp
97	7	0.9	249	17	Q9YEH3	Q9yeh3 aeropyrum p	170	7	0.9	394	16	O51791	O51791 borrelia bu
98	7	0.9	250	17	O27347	O27347 methanobact	171	7	0.9	394	16	Q8EEC2	Q8eec2 shewanella
99	7	0.9	251	5	O16973	O16973 caenorhabdi	172	7	0.9	395	2	Q8KTP4	Q8ktp4 candidatus
100	7	0.9	251	16	Q8ZH32	Q8zh32 versinia pe	173	7	0.9	398	16	Q8XJH1	Q8xjh1 clostridium
101	7	0.9	262	16	Q8Y3G0	Q8y3g0 kalsonia s	174	7	0.9	398	16	Q8P5B8	Q8p5b8 xanthomonas
102	7	0.9	263	16	Q9WZC8	Q9wzc8 thermotoga	175	7	0.9	399	16	Q8CV34	Q8cv34 oceanobacil
103	7	0.9	265	2	Q8LZD5	Q8lzd5 proteus vul	176	7	0.9	401	5	O18148	O18148 caenorhabdi
104	7	0.9	265	16	Q9RKR8	Q9rkr8 streptomyce	177	7	0.9	402	16	Q8PPI9	Q8ppi9 xanthomonas
105	7	0.9	266	4	Q9H0J0	Q9hoj0 homo sapien	178	7	0.9	403	16	Q9ABQ1	Q9abq1 caulobacter
106	7	0.9	267	8	Q94ZQ4	Q94zq4 myotis myot	179	7	0.9	405	16	Q8NZ75	Q8nzt75 streptococc
107	7	0.9	267	8	Q94VT8	Q94vt8 eptesicus b	180	7	0.9	405	16	Q99XU4	Q99xu4 streptococc
108	7	0.9	267	8	Q94VW8	Q94vw8 myotis myot	181	7	0.9	406	16	O66947	O66947 aquifex seo
109	7	0.9	267	8	Q94ZQ2	Q94zq2 myotis blyt	182	7	0.9	407	5	Q8MWG4	Q8mwg4 patella vul
110	7	0.9	267	8	Q94VW8	Q94vw8 myotis myot	183	7	0.9	410	10	Q8GSI8	Q8gsi8 oryza sativ
111	7	0.9	267	8	Q94VT9	Q94vt9 eptesicus b	184	7	0.9	413	10	Q94H79	Q94h79 oryza sativ
112	7	0.9	267	8	Q94VX1	Q94vx1 myotis myot	185	7	0.9	414	16	Q8UDV4	Q8udv4 agrobacteri
113	7	0.9	267	8	Q94VW7	Q94vw7 myotis blyt	186	7	0.9	415	5	Q8T745	Q8t745 branchiost
114	7	0.9	267	8	Q94VX0	Q94vx0 myotis myot	187	7	0.9	415	16	O06177	O06177 mycobacteri
115	7	0.9	268	5	O76290	O76290 trypanosoma	188	7	0.9	416	2	O85164	O85164 nitrosococc
116	7	0.9	273	17	Q978L6	Q978l6 thermoplasma	189	7	0.9	417	16	Q8DLH5	Q8dlh5 synechococc
117	7	0.9	276	10	Q9SEW3	Q9sew3 glycine max	190	7	0.9	418	11	Q8CI40	Q8ci40 mus musculu
118	7	0.9	278	2	Q8GQAI	Q8gqai pseudomonas	191	7	0.9	418	12	Q9QCUE	Q9qcu6 odocoileus
119	7	0.9	279	16	Q8DM47	Q8dm47 synechococc	192	7	0.9	419	2	Q9PAI0	Q9pai0 nitrosococc
120	7	0.9	280	16	Q8CZ23	Q8cz23 yersinia pe	193	7	0.9	423	17	Q8Q0W7	Q8q0m7 methanosarc
121	7	0.9	280	17	Q8L172	Q8l172 archaeoglob	194	7	0.9	428	2	O69000	O69000 mycobacteri
122	7	0.9	285	17	Q9HMH8	Q9hnm8 halobacteri	195	7	0.9	428	2	Q8GEA2	Q8gea2 mycobacteri
123	7	0.9	285	17	Q9P8B0	Q9psb0 methanosarc	196	7	0.9	430	16	Q8G250	Q8g250 bruceella su
124	7	0.9	286	16	Q9KYQ0	Q9kyq0 streptomyce	197	7	0.9	434	16	Q9WZ39	Q9wz99 thermotoga
125	7	0.9	290	11	Q8BX33	Q8bx33 mus musculu	198	7	0.9	434	16	Q8NLQ2	Q8nlq2 corynebacte
126	7	0.9	290	11	Q8BMD4	Q8bmd4 mus musculu	199	7	0.9	437	5	Q19304	Q19304 caenorhabdi
127	7	0.9	292	10	Q94ZU1	Q94zul oryza sativ	200	7	0.9	441	16	Q8D384	Q8d384 wigglsworth
128	7	0.9	294	5	Q9XWY8	Q9xwy8 caenorhabdi	201	7	0.9	442	10	Q9C8I6	Q9c8y6 arabidopsis
129	7	0.9	296	11	Q921C7	Q921c7 mus caroli	202	7	0.9	443	5	Q95QH7	Q95qht7 caenorhabdi
130	7	0.9	302	5	Q9VCJ0	Q9vcj0 drosophila	203	7	0.9	443	10	O23013	O23013 arabidopsis
131	7	0.9	303	5	Q22674	Q22674 caenorhabdi	204	7	0.9	444	2	Q56568	Q56568 vibrio angu
132	7	0.9	303	11	Q91X08	Q91x08 mus musculu	205	7	0.9	444	16	Q8XKX5	Q8xxk5 clostridium
133	7	0.9	309	16	Q8CU42	Q8cu42 staphylococc	206	7	0.9	448	10	O81452	O81452 arabidopsis
134	7	0.9	310	2	Q9RCU3	Q9rcj3 streptococc	207	7	0.9	453	16	Q92JY3	Q92jy3 rhizobium m
135	7	0.9	310	16	Q9KB56	Q9kb56 bacillus ha	208	7	0.9	455	16	Q92S78	Q92s78 rhizobium m
136	7	0.9	312	16	Q9CBP7	Q9cbp7 mycobacteri	209	7	0.9	458	17	Q8PT72	Q8pt72 methanosarc
137	7	0.9	313	10	Q94E82	Q94e82 oryza sativ	210	7	0.9	459	10	Q94GT1	Q94gt1 oryza sativ
138	7	0.9	316	12	O57092	O57092 ectromelia	211	7	0.9	460	8	Q94S81	Q94s81 platichthys
139	7	0.9	316	16	Q9PND2	Q9pnd2 campylobact	212	7	0.9	466	3	Q9E6Y5	Q9pey5 neurospora
140	7	0.9	318	8	Q8M884	Q8m884 piecotus au	213	7	0.9	466	11	Q8R1D7	Q8rid7 mus musculu
141	7	0.9	318	11	Q8CCA5	Q8cca5 mus musculu	214	7	0.9	467	10	Q8H8X6	Q8h8x6 oryza sativ
142	7	0.9	319	2	Q48429	Q48429 klebsiella	215	7	0.9	468	5	Q20446	Q20446 caenorhabdi
143	7	0.9	320	5	Q9UA62	Q9ua62 caenorhabdi	216	7	0.9	468	16	Q9K0K1	Q9k0k1 neisseria m
144	7	0.9	320	12	O57079	O57079 cowpox viru	217	7	0.9	468	16	Q9JVK7	Q9jvk7 neisseria m
145	7	0.9	320	12	O57300	O57300 ectromelia	218	7	0.9	469	10	Q9PXF6	Q9pxf6 arabidopsis
146	7	0.9	321	4	Q8NGK1	Q8ngk1 homo sapien	219	7	0.9	469	16	Q9HTL2	Q9htl2 pseudomonas
147	7	0.9	322	11	Q9JMH0	Q9jmh0 rattus norv	220	7	0.9	470	5	Q9VIQ2	Q9viq2 drosophila
148	7	0.9	322	12	O72761	O72761 cowpox viru	221	7	0.9	470	16	Q9YFQ2	Q9yfq2 bruceella me
149	7	0.9	325	11	Q9EP88	Q9ep88 rattus norv	222	7	0.9	470	16	Q8G266	Q8g266 bruceella su
150	7	0.9	327	5	Q9XXB0	Q9xxb0 caenorhabdi	223	7	0.9	475	2	Q59114	Q59114 acetomicrob
151	7	0.9	331	16	Q984X1	Q984x1 rhizobium 1	224	7	0.9	476	11	Q91ZN7	Q91zn7 rattus norv
152	7	0.9	332	2	Q9FIV1	Q9fiv1 kitasatospo	225	7	0.9	476	11	Q9D0N2	Q9don2 mus musculu
153	7	0.9	335	15	O41896	O41896 bovine foam	226	7	0.9	478	16	Q92M49	Q92m49 rhizobium m
154	7	0.9	335	15	Q8ALT8	Q8alt8 bovine foam	227	7	0.9	485	5	Q9W4L8	Q9w4l8 drosophila
155	7	0.9	338	16	Q9RKN6	Q9rkn6 streptomyce	228	7	0.9	488	16	Q92C04	Q92c04 listeria in
156	7	0.9	338	17	Q9HQP7	Q9hqp7 halobacteri	229	7	0.9	488	16	Q8Y7D3	Q8y7d3 listeria mo
157	7	0.9	343	5	Q9XXS2	Q9xxs2 caenorhabdi	230	7	0.9	490	16	Q8IFR9	Q8ifr9 bruceella me
158	7	0.9	346	16	Q8PJT4	Q8pjt4 xanthomonas	231	7	0.9	492	10	O80830	O80830 arabidopsis
159	7	0.9	353	10	Q8LEY4	Q8ley4 arabidopsis	232	7	0.9	496	16	Q8CJPS	Q8cjp5 streptomyce
160	7	0.9	353	10	Q9FR28	Q9fr28 petroselinu	233	7	0.9	497	2	Q30986	Q30986 staphylococ
161	7	0.9	353	10	Q9LYZ6	Q9lyz6 arabidopsis	234	7	0.9	497	2	Q53584	Q53584 staphylococ
162	7	0.9	356	16	Q8G649	Q8g649 bifidobacte	235	7	0.9	498	16	Q9A9H5	Q9a9h5 caulobacter

236	7	0.9	501	4	Q8IVG2	Q8ivg2 homo sapien	309	7	0.9	765	2	Q8KMS6	Q8kms6 flavobacter
237	7	0.9	502	2	Q8RHR4	Q8rnhr4 streptomyce	310	7	0.9	770	5	Q9XWP6	Q9xwp6 caenorhabdi
238	7	0.9	504	12	Q8JU09	Q8juu9 eucurbit ye	311	7	0.9	774	5	O46021	O46021 caenorhabdi
239	7	0.9	507	10	Q9LHV2	Q9lhy2 cryza sativ	312	7	0.9	776	5	Q8I4A7	Q8i4a7 caenorhabdi
240	7	0.9	509	2	Q8KWA2	Q8kwa2 ruegeria sp	313	7	0.9	781	17	Q8PT66	Q8pt66 methanosarc
241	7	0.9	509	10	Q9C9J6	Q9c9j6 arabidopsis	314	7	0.9	782	17	Q97VPS	Q97vr5 sulfolobus
242	7	0.9	512	16	Q99SV5	Q99sv5 staphylococ	315	7	0.9	791	5	Q8WQ96	Q8wq96 caenorhabdi
243	7	0.9	513	16	Q9X9U1	Q9x9u1 streptomyce	316	7	0.9	791	16	Q8CUG5	Q8cug5 oceanobacil
244	7	0.9	520	10	Q9SR68	Q9sr68 arabidopsis	317	7	0.9	797	5	Q9VQN3	Q9vqn3 drosophila
245	7	0.9	524	16	O53294	O53294 mycobacteri	318	7	0.9	800	5	O96902	O96902 dictyosteli
246	7	0.9	527	10	Q8W4R7	Q8w4r7 arabidopsis	319	7	0.9	802	16	Q92C16	Q92c16 listeria in
247	7	0.9	533	5	Q9W520	Q9w520 drosophila	320	7	0.9	802	16	Q8Y7Q1	Q8y7q1 listeria mo
248	7	0.9	533	5	O46096	O46096 drosophila	321	7	0.9	806	5	Q8ICH6	Q8ich6 plasmodium
249	7	0.9	537	3	O13369	O13369 ascobolus i	322	7	0.9	843	10	Q8L7W8	Q8l7w8 arabidopsis
250	7	0.9	537	10	Q9SL89	Q9sl89 arabidopsis	323	7	0.9	844	12	Q8UJZ9	Q8ujz9 infectious
251	7	0.9	539	10	Q19079	Q19079 rhizobium l	324	7	0.9	844	12	Q8UJZ7	Q8ujz7 infectious
252	7	0.9	539	16	Q988C2	Q988c2 rhizobium l	325	7	0.9	847	10	O9SYZ1	O9syz1 arabidopsis
253	7	0.9	542	10	Q94JR1	Q94jr1 arabidopsis	326	7	0.9	850	5	Q8NA22	Q8na22 caenorhabdi
254	7	0.9	546	10	Q9C5H2	Q9c5h2 arabidopsis	327	7	0.9	852	4	Q8NV8	Q8nv8 homo sapien
255	7	0.9	549	16	Q8KGB4	Q8kgb4 chlorobium	328	7	0.9	852	4	Q8NEH2	Q8neh2 homo sapien
256	7	0.9	554	2	Q9JTE6	Q9jete6 rabnella aq	329	7	0.9	852	16	Q8FM94	Q8fm94 corynebacte
257	7	0.9	555	16	Q9HUX8	Q9hux8 pseudomonas	330	7	0.9	892	16	Q92AT7	Q92at7 listeria in
258	7	0.9	556	13	Q9PWA2	Q9pwa2 gallus gall	331	7	0.9	892	16	Q8F846	Q8f846 leptospira
259	7	0.9	560	2	Q9KGX9	Q9kgx9 mycoplasma	332	7	0.9	910	10	Q8H912	Q8h912 cryza sativ
260	7	0.9	560	2	Q9KGX7	Q9kgx7 mycoplasma	333	7	0.9	930	5	Q8WQX9	Q8wqx9 crypanosoma
261	7	0.9	561	17	Q97A25	Q97az5 thermoplasma	334	7	0.9	937	16	Q8PPY9	Q8ppy9 xanthomonas
262	7	0.9	565	10	Q8LSQ3	Q8lsq3 cryza sativ	335	7	0.9	956	5	O00908	O00908 cryptospori
263	7	0.9	569	2	Q8RFF2	Q8rfi2 rhodobacter	336	7	0.9	958	16	Q8EBS5	Q8eb55 shewanella
264	7	0.9	576	16	Q8R9B5	Q8r9b5 thermoanaer	337	7	0.9	966	16	Q9K7X8	Q9k7x8 bacillus ha
265	7	0.9	577	16	Q8F4D6	Q8f4d6 leptospira	338	7	0.9	970	5	Q9VN28	Q9vn28 drosophila
266	7	0.9	580	10	Q9LMV8	Q9lmv8 arabidopsis	339	7	0.9	970	16	Q8DY17	Q8dy17 streptococc
267	7	0.9	587	5	O96459	O96459 strongyloce	340	7	0.9	990	15	O41895	O41895 bovine foam
268	7	0.9	587	5	Q8L1Y0	Q8l1y0 aeromonas h	341	7	0.9	990	15	Q8ALU0	Q8alu0 bovine foam
269	7	0.9	594	2	Q8RT82	Q8rt82 aeromonas h	342	7	0.9	1003	15	Q8UT43	Q8ut43 human immun
270	7	0.9	594	10	Q9CTV9	Q9ctv9 arabidopsis	343	7	0.9	1018	16	Q9HW27	Q9hw27 pseudomonas
271	7	0.9	599	17	Q8TX47	Q8tx47 methanopyru	344	7	0.9	1043	5	Q9VL84	Q9vl84 drosophila
272	7	0.9	601	17	Q9V0A2	Q9v0a2 pyrococcus	345	7	0.9	1046	16	Q9AAN5	Q9aan5 caulobacter
273	7	0.9	605	10	Q9LS19	Q9ls19 arabidopsis	346	7	0.9	1052	2	Q93NP0	Q93np0 actinobacil
274	7	0.9	623	16	Q9CBP6	Q9cbp6 mycobacteri	347	7	0.9	1067	13	Q8AVK8	Q8avk8 xenopus lae
275	7	0.9	626	10	Q9FKY0	Q9fky0 arabidopsis	348	7	0.9	1070	4	O95206	O95206 homo sapien
276	7	0.9	627	5	Q93212	Q93212 caenorhabdi	349	7	0.9	1070	4	O8IYE9	O8ive9 homo sapien
277	7	0.9	628	16	Q9A4M9	Q9a4m9 caulobacter	350	7	0.9	1071	11	Q9JKP3	Q9jkp3 mus musculu
278	7	0.9	629	5	Q8IQ02	Q8iq02 drosophila	351	7	0.9	1073	5	Q9VRA3	Q9vra3 drosophila
279	7	0.9	642	16	Q9A312	Q9a312 caulobacter	352	7	0.9	1084	16	Q8KFL5	Q8kfl5 chlorobium
280	7	0.9	645	16	Q8XDL8	Q8xdl8 escherichia	353	7	0.9	1086	5	Q9TYN3	Q9tyn3 caenorhabdi
281	7	0.9	645	16	Q8X384	Q8x384 escherichia	354	7	0.9	1148	5	Q8IQ31	Q8iq31 drosophila
282	7	0.9	648	5	Q8P152	Q8p152 drosophila	355	7	0.9	1186	16	O8YV73	O8yv73 anabaena sp
283	7	0.9	648	16	Q8P154	Q8p154 xanthomonas	356	7	0.9	1188	10	O9FKN9	O9fkn9 arabidopsis
284	7	0.9	658	17	Q9V2Q2	Q9v2q2 pyrococcus	357	7	0.9	1252	16	Q97GV3	Q97gv3 clostridium
285	7	0.9	663	2	Q8RR60	Q8rr60 rhizobium s	358	7	0.9	1310	16	Q8E473	Q8e473 streptococc
286	7	0.9	666	4	O8N8G9	O8n8g9 homo sapien	359	7	0.9	1313	4	O75033	O75033 homo sapien
287	7	0.9	669	8	O8R477	O8r477 desfontaini	360	7	0.9	1342	5	Q9VXG6	Q9vxg6 drosophila
288	7	0.9	677	3	Q8X0T5	Q8x0t5 neurospora	361	7	0.9	1366	16	O92529	O92529 streptomyce
289	7	0.9	680	4	Q96AC1	Q96ac1 homo sapien	362	7	0.9	1367	11	Q9EQD9	Q9eqd9 mus musculu
290	7	0.9	680	11	Q8CIB5	Q8cib5 mus musculu	363	7	0.9	1367	11	O35206	O35206 mus musculu
291	7	0.9	698	13	Q9PU71	Q9pu71 xenopus lae	364	7	0.9	1400	3	Q9HFK8	Q9hfk8 neurospora
292	7	0.9	704	16	Q8KC51	Q8kc51 chlorobium	365	7	0.9	1464	12	Q8UZ08	Q8uz08 rice black
293	7	0.9	707	10	Q9FGU4	Q9fgu4 arabidopsis	366	7	0.9	1464	12	O8B4V7	O8b4v7 rice black
294	7	0.9	707	10	Q9ARN9	Q9arn9 cryza sativ	367	7	0.9	1469	10	Q9LKR1	Q9lkr1 pisum sativ
295	7	0.9	713	11	Q9D2Q2	Q9d2q2 mus musculu	368	7	0.9	1734	4	Q8RE24	Q8re24 homo sapien
296	7	0.9	720	4	Q14840	Q14840 homo sapien	369	7	0.9	1832	5	O96503	O96503 cryptospori
297	7	0.9	724	5	Q9GTJ8	Q9gtj8 hydra atten	370	7	0.9	1968	12	Q9WA89	Q9wa89 viral hemor
298	7	0.9	728	16	Q8D719	Q8d719 vibrio vuln	371	7	0.9	1984	12	Q9QBH5	Q9qbh5 viral hemor
299	7	0.9	729	2	Q8Q415	Q8q415 pseudomonas	372	7	0.9	1984	12	O9QJY4	O9qjy4 viral hemor
300	7	0.9	729	5	Q17796	Q17796 caenorhabdi	373	7	0.9	1984	12	O9YQ45	O9yq45 viral hemor
301	7	0.9	736	10	Q8L7D2	Q8l7d2 arabidopsis	374	7	0.9	1984	12	Q9QSW0	Q9qsw0 viral hemor
302	7	0.9	737	16	O8YVW9	O8yvww9 anabaena sp	375	7	0.9	1984	12	O9QBH1	O9qbh1 viral hemor
303	7	0.9	741	8	O98703	O98703 desfontaini	376	7	0.9	1987	12	O9WME7	O9wme7 viral hemor
304	7	0.9	756	2	O9X5P3	O9x5p3 pseudomonas	377	7	0.9	2009	11	O8CJ40	O8cj40 mus musculu
305	7	0.9	760	10	Q9SDY2	Q9sdy2 nicotiana t	378	7	0.9	2089	11	Q39478	Q39478 cyclocella
306	7	0.9	763	16	Q8XJY4	Q8xjy4 clostridium	379	7	0.9	2126	2	Q9RNB0	Q9rnrb0 microcystis
307	7	0.9	764	2	O68881	O68881 pseudomonas	380	7	0.9				
308	7	0.9	764	16	Q9HV88	Q9hv88 pseudomonas	381	7	0.9				

382	7	0.9	2142	10	Q9FWR3	Q9fwr3 arabidopsis	455	6	0.8	62	16	Q5102	O25102 helicobacte
383	7	0.9	2283	2	Q8VQ99	Q8vq99 staphylococ	456	6	0.8	62	16	Q99UF3	Q99uf3 staphylococ
384	7	0.9	3263	16	Q8VQ197	Q8vq197 vibrio chol	457	6	0.8	63	10	Q8RZZ2	Q8rzz2 oryza sativ
385	7	0.9	3419	11	O55147	O55147 rattus norv	458	6	0.8	64	10	Q8W0J3	Q8w0j3 oryza sativ
386	7	0.9	3429	11	O08614	O08614 mus musculu	459	6	0.8	64	16	Q9C116	Q9c116 lactococcus
387	7	0.9	3638	4	Q15142	Q15142 homo sapien	460	6	0.8	65	8	Q9MUH6	Q9muh6 pteridium a
388	7	0.9	3906	2	Q9FNU1	Q9fdu1 microcystis	461	6	0.8	66	16	Q99R11	Q99r11 staphylococ
389	7	0.9	3908	5	Q9BK91	Q9bk91 strongyloce	462	6	0.8	66	16	Q8NUL3	Q8nul3 staphylococ
390	7	0.9	4256	6	Q8MJF3	Q8mjf3 canis famil	463	6	0.8	67	8	Q9MU72	Q9mu72 pteridium a
391	7	0.9	4283	11	Q9ERV0	Q9erv0 rattus norv	464	6	0.8	67	8	Q9MUH3	Q9muh3 pteridium e
392	7	0.9	4292	4	Q15141	Q15141 homo sapien	465	6	0.8	67	8	Q9MUH2	Q9muh2 pteridium a
393	7	0.9	4293	11	O08852	O08852 mus musculu	466	6	0.8	67	10	Q8GT05	Q8gt05 oryza sativ
394	7	0.9	4302	4	Q15140	Q15140 homo sapien	467	6	0.8	67	12	Q9E1J1	Q9e1j1 sin nombre
395	7	0.9	5938	4	Q96PK2	Q96pk2 homo sapien	468	6	0.8	67	17	Q973Z1	Q973z1 sulfolobus
396	6	0.8	16	2	Q9RAJ4	Q9raj4 pseudomonas	469	6	0.8	68	2	Q9ZFK9	Q9zfk9 myxococcus
397	6	0.8	25	12	Q9QC21	Q9qc21 hepatitis c	470	6	0.8	68	8	Q9MUH5	Q9muh5 pteridium a
398	6	0.8	27	3	Q9URB3	Q9urb3 candida alb	471	6	0.8	68	16	Q8VK99	Q8vk99 mycobacteri
399	6	0.8	27	12	Q91LD4	Q91ld4 hepatitis c	472	6	0.8	68	16	Q8VKA6	Q8vka6 fusobacteri
400	6	0.8	27	13	O57558	O57558 lampetra pl	473	6	0.8	69	16	Q8RIA4	Q8ria4 corynebacte
401	6	0.8	28	12	Q9QC38	Q9qc38 hepatitis c	474	6	0.8	70	2	Q9FB43	Q9fb43 corynebacte
402	6	0.8	28	12	Q9QC28	Q9qc28 hepatitis c	475	6	0.8	70	10	Q9LLK5	Q9llk5 pyrus pyrif
403	6	0.8	28	12	Q9QC25	Q9qc25 hepatitis c	476	6	0.8	70	12	Q8UX25	Q8ux25 hepatitis c
404	6	0.8	28	12	Q9QC44	Q9qc44 hepatitis c	477	6	0.8	71	5	Q9UAA6	Q9uaa6 conus livid
405	6	0.8	28	12	Q9QC43	Q9qc43 hepatitis c	478	6	0.8	71	16	Q8UB92	Q8ub92 agrobacteri
406	6	0.8	28	12	Q9QC22	Q9qc22 hepatitis c	479	6	0.8	72	16	Q8UB15	Q8ub15 agrobacteri
407	6	0.8	28	12	Q9QC39	Q9qc39 hepatitis c	480	6	0.8	73	16	Q88801	Q88801 pseudomonas
408	6	0.8	28	12	Q9QC37	Q9qc37 hepatitis c	481	6	0.8	73	16	Q91I29	Q91i29 pseudomonas
409	6	0.8	28	12	Q9QC41	Q9qc41 hepatitis c	482	6	0.8	73	16	Q8FN91	Q8fn91 corynebacte
410	6	0.8	28	12	Q9QC34	Q9qc34 hepatitis c	483	6	0.8	73	16	Q8ERF6	Q8erf6 oceanobacil
411	6	0.8	28	12	Q9QC31	Q9qc31 hepatitis c	484	6	0.8	75	16	Q8UJ18	Q8uj18 agrobacteri
412	6	0.8	28	12	Q9QC40	Q9qc40 hepatitis c	485	6	0.8	76	5	Q8WS66	Q8ws66 glossina mo
413	6	0.8	28	12	Q9QC42	Q9qc42 hepatitis c	486	6	0.8	77	12	Q9E1J4	Q9e1j4 sin nombre
414	6	0.8	28	12	Q9QC29	Q9qc29 hepatitis c	487	6	0.8	77	17	Q979C5	Q979c5 thermoplasma
415	6	0.8	28	12	Q9QC23	Q9qc23 hepatitis c	488	6	0.8	78	15	Q70471	Q70471 human immun
416	6	0.8	28	12	Q9QC30	Q9qc30 hepatitis c	489	6	0.8	81	3	Q92461	Q92461 yarrowia li
417	6	0.8	28	12	Q9QC27	Q9qc27 hepatitis c	490	6	0.8	81	3	O59839	O59839 candida alb
418	6	0.8	28	12	Q9QC33	Q9qc33 hepatitis c	491	6	0.8	81	7	O19462	O19462 mus musculu
419	6	0.8	28	12	Q9QC14	Q9qc14 hepatitis c	492	6	0.8	81	7	Q9XRP3	Q9xrp3 mus spicile
420	6	0.8	28	12	Q9QC20	Q9qc20 hepatitis c	493	6	0.8	81	7	Q9XRN9	Q9xrn9 mus musculu
421	6	0.8	28	12	Q9QC36	Q9qc36 hepatitis c	494	6	0.8	81	7	Q9XRP2	Q9xrp2 mus macedon
422	6	0.8	28	12	Q9QC45	Q9qc45 hepatitis c	495	6	0.8	81	7	Q9XRP1	Q9xrp1 mus spretus
423	6	0.8	28	12	Q9QC15	Q9qc15 hepatitis c	496	6	0.8	81	7	Q9XRP0	Q9xrp0 mus musculu
424	6	0.8	28	12	Q9QC32	Q9qc32 hepatitis c	497	6	0.8	81	10	Q94D02	Q94d02 oryza sativ
425	6	0.8	28	12	Q9QC26	Q9qc26 hepatitis c	498	6	0.8	81	12	Q9DPS3	Q9dps3 meleagrid h
426	6	0.8	32	5	Q17070	Q17070 antheraea p	499	6	0.8	81	12	Q9E1H9	Q9e1h9 meleagrid h
427	6	0.8	36	8	Q9TFV4	Q9tfv4 teratoscinc	500	6	0.8	81	12	Q9E1H9	Q9e1h9 meleagrid h
428	6	0.8	36	8	Q9TFV1	Q9tfv1 teratoscinc	501	6	0.8	81	16	Q8NM13	Q8nm13 corynebacte
429	6	0.8	37	8	Q8M6Z4	Q8m6z4 alle alle (502	6	0.8	81	16	Q8DM70	Q8dm70 synchococcc
430	6	0.8	46	16	Q8P1S5	Q8p1s5 streptococ	503	6	0.8	82	12	Q9Q3S8	Q9q3s8 hepatitis c
431	6	0.8	46	16	Q8P082	Q8p082 streptococ	504	6	0.8	82	16	Q97RQ1	Q97rq1 streptococ
432	6	0.8	46	16	Q8E104	Q8e104 streptococ	505	6	0.8	82	16	Q8DQ12	Q8dq12 streptococ
433	6	0.8	47	5	Q8W1I3	Q8wti3 blackburnia	506	6	0.8	83	5	Q81FHS	Q81fh5 trypanosoma
434	6	0.8	47	16	Q8EXA0	Q8exa0 leptospira	507	6	0.8	83	5	Q81JH2	Q81jh2 plasmodium
435	6	0.8	50	12	Q9DQ83	Q9dg83 pineapple m	508	6	0.8	85	10	Q99197	Q99197 chlamydomon
436	6	0.8	50	16	Q8UES0	Q8ues0 agrobacteri	509	6	0.8	85	10	Q8H6H6	Q8h6h6 triticum ae
437	6	0.8	51	5	O43364	O43364 ceratitis c	510	6	0.8	85	16	Q9PBF0	Q9pbf0 xyella fas
438	6	0.8	51	8	Q36043	Q36043 tetragma ge	511	6	0.8	86	12	Q9QN76	Q9qn76 ectocarpus
439	6	0.8	54	8	Q957R3	Q957r3 dogania sub	512	6	0.8	87	3	Q9UW72	Q9uw72 cryptococcc
440	6	0.8	56	8	Q9MUH4	Q9muh4 pteridium a	513	6	0.8	87	10	O81353	O81353 gosepyium h
441	6	0.8	56	8	Q9MUH1	Q9muh1 pteridium a	514	6	0.8	88	3	Q9UW71	Q9uw71 cryptococcc
442	6	0.8	57	16	Q931Y3	Q931y3 staphylococ	515	6	0.8	88	5	Q8MLC0	Q8mlc0 drosophila
443	6	0.8	57	8	Q9MUH9	Q9muh9 pteridium a	516	6	0.8	88	10	Q94AZ8	Q94az8 arabidopsis
444	6	0.8	57	12	Q64802	Q64802 autographa	517	6	0.8	89	2	Q9REP2	Q9rep2 zymomonas m
445	6	0.8	57	16	Q97PPI	Q97ppi streptococ	518	6	0.8	89	3	Q9UW74	Q9uw74 cryptococcc
446	6	0.8	57	16	Q8CYI4	Q8cyi4 streptococ	519	6	0.8	89	5	Q815X2	Q815x2 plasmodium
447	6	0.8	58	8	Q9MU10	Q9mu10 pteridium a	520	6	0.8	89	7	O62866	O62866 rattus norv
448	6	0.8	58	8	Q9MU71	Q9mu71 pteridium a	521	6	0.8	89	7	O62865	O62865 rattus norv
449	6	0.8	60	16	Q9K8V9	Q9k8v9 bacillus ha	522	6	0.8	89	8	Q95888	Q95888 peromyscus
450	6	0.8	60	2	Q8VSN6	Q8vsn6 shigella fl	523	6	0.8	89	10	O81354	O81354 gossypium h
451	6	0.8	60	8	Q9MUH8	Q9muh8 pteridium p	524	6	0.8	89	16	Q9X0Q8	Q9x0q8 thermotoga
452	6	0.8	61	2	O54485	O54485 yersinia pe	525	6	0.8	89	16	Q8DJT4	Q8dj14 synechococ
453	6	0.8	61	2	O54485	O54485 staphylococ	526	6	0.8	91	3	Q9UW77	Q9uw77 cryptococcc
454	6	0.8	62	3	Q96WL5	Q96wl5 aspergillus	527	6	0.8	91	3	Q9UW73	Q9uw73 cryptococcc

528	6	0.8	91	10	Q8H0Y6	Q8h0y6 arabidopsis	601	5	Q95ZJ6	Q95zj6 caenorhabdi
529	6	0.8	91	16	Q8F6S6	Q8f6s6 leptospira	602	10	Q93XD0	Q93xd0 zea mays (m
530	6	0.8	92	3	Q9UW70	Q9uw70 cryptococcu	603	11	Q9CX18	Q9cx18 mus musculu
531	6	0.8	92	10	Q8L7G7	Q8l7g7 arabidopsis	604	10	O11343	O11343 molluscum c
532	6	0.8	92	10	Q8GW21	Q8gw21 arabidopsis	605	16	Q97HT4	Q97ht4 clostridium
533	6	0.8	92	12	Q86604	Q86604 rembrandt t	606	16	O8F117	O8f117 leptospira
534	6	0.8	93	3	Q9UQ26	Q9uq26 cryptococcu	607	16	O8ECM4	O8ecm4 shewanella
535	6	0.8	93	3	Q9UW79	Q9uw79 cryptococcu	608	16	O8ECM4	O8ecm4 shewanella
536	6	0.8	93	3	Q9UQ27	Q9uq27 cryptococcu	609	5	Q9NLB2	Q9nlb2 plasmodium
537	6	0.8	93	16	Q8X1U2	Q8xiu2 clostridium	610	12	O11782	O11782 puumala vir
538	6	0.8	94	3	Q9UW78	Q9uw78 cryptococcu	611	16	Q99ZA1	Q99za1 streptococc
539	6	0.8	94	3	Q9UR31	Q9ur31 cryptococcu	612	16	O8P0J2	O8p0j2 streptococc
540	6	0.8	94	3	Q9UR32	Q9ur32 cryptococcu	613	16	O8K738	O8k738 streptococc
541	6	0.8	94	10	O8L951	Q8l951 arabidopsis	614	16	O8L951	Q8l951 arabidopsis
542	6	0.8	94	16	Q9KN60	Q9kn60 vibrio chol	615	16	O8L951	Q8l951 arabidopsis
543	6	0.8	95	6	Q9BF30	Q9bf30 procavia ca	616	16	O8L951	Q8l951 arabidopsis
544	6	0.8	95	16	Q92NX8	Q92nx8 rhizobium m	617	16	O8L951	Q8l951 arabidopsis
545	6	0.8	96	2	Q9F523	Q9f523 escherichia	618	16	O8L951	Q8l951 arabidopsis
546	6	0.8	96	10	O8H7T6	Q8h7t6 oryza sativ	619	16	O8L951	Q8l951 arabidopsis
547	6	0.8	96	16	O8F6T4	Q8f6t4 leptospira	620	16	O8L951	Q8l951 arabidopsis
548	6	0.8	97	12	O8F3I4	Q8f3i4 maryland ca	621	16	O8L951	Q8l951 arabidopsis
549	6	0.8	97	16	Q98M37	Q98m37 rhizobium l	622	16	O8L951	Q8l951 arabidopsis
550	6	0.8	97	16	Q8P2G8	Q8p2g8 streptococc	623	16	O8L951	Q8l951 arabidopsis
551	6	0.8	98	2	Q9JPF9	Q9jpf9 neisseria m	624	16	O8L951	Q8l951 arabidopsis
552	6	0.8	98	2	Q9JPD6	Q9jpd6 neisseria m	625	16	O8L951	Q8l951 arabidopsis
553	6	0.8	98	4	Q9NDT3	Q9ndt3 homo sapien	626	16	O8L951	Q8l951 arabidopsis
554	6	0.8	98	16	Q988W9	Q988w9 rhizobium l	627	16	O8L951	Q8l951 arabidopsis
555	6	0.8	99	12	Q9L1U7	Q9liu7 influenza a	628	16	O8L951	Q8l951 arabidopsis
556	6	0.8	100	2	O52301	O52301 escherichia	629	16	O8L951	Q8l951 arabidopsis
557	6	0.8	100	2	O52303	O52303 escherichia	630	16	O8L951	Q8l951 arabidopsis
558	6	0.8	100	3	Q9UT62	Q9ut62 schizosacch	631	16	O8L951	Q8l951 arabidopsis
559	6	0.8	100	10	Q9AXT5	Q9axt5 brassica ca	632	16	O8L951	Q8l951 arabidopsis
560	6	0.8	100	16	Q9PNL7	Q9pnl7 campylobact	633	16	O8L951	Q8l951 arabidopsis
561	6	0.8	100	16	O8ZHX2	Q8zhx2 yersinia pe	634	16	O8L951	Q8l951 arabidopsis
562	6	0.8	100	16	O8YRX3	Q8yrx3 anabaena sp	635	16	O8L951	Q8l951 arabidopsis
563	6	0.8	100	17	Q97BHS	Q97bhs thermoplasm	636	16	O8L951	Q8l951 arabidopsis
564	6	0.8	100	17	O8ZXN9	Q8zxm9 pyrobaculum	637	16	O8L951	Q8l951 arabidopsis
565	6	0.8	101	2	Q9ETD9	Q9etd9 bdellovibri	638	16	O8L951	Q8l951 arabidopsis
566	6	0.8	101	5	O8WS43	Q8ws43 oikopleura	639	16	O8L951	Q8l951 arabidopsis
567	6	0.8	101	9	Q8W6V5	Q8w6v5 cyanophage	640	16	O8L951	Q8l951 arabidopsis
568	6	0.8	101	16	Q9HZ85	Q9hz85 pseudomonas	641	16	O8L951	Q8l951 arabidopsis
569	6	0.8	102	13	Q9L811	Q9l811 xenopus lae	642	16	O8L951	Q8l951 arabidopsis
570	6	0.8	103	2	Q46697	Q46697 escherichia	643	16	O8L951	Q8l951 arabidopsis
571	6	0.8	104	4	O96QC3	Q96qc3 homo sapien	644	16	O8L951	Q8l951 arabidopsis
572	6	0.8	104	4	O8WY72	Q8wy72 homo sapien	645	16	O8L951	Q8l951 arabidopsis
573	6	0.8	104	5	Q9U601	Q9u601 anisakis si	646	16	O8L951	Q8l951 arabidopsis
574	6	0.8	104	9	Q37864	Q37864 bacterioph	647	16	O8L951	Q8l951 arabidopsis
575	6	0.8	104	9	O21876	O21876 bacterioph	648	16	O8L951	Q8l951 arabidopsis
576	6	0.8	104	9	O80109	O80109 bacterioph	649	16	O8L951	Q8l951 arabidopsis
577	6	0.8	104	16	O9CLF6	Q9clf6 pasteurella	650	16	O8L951	Q8l951 arabidopsis
578	6	0.8	104	16	Q97NN4	Q97nn4 streptococc	651	16	O8L951	Q8l951 arabidopsis
579	6	0.8	104	16	O9FBM5	Q9fbm5 streptococc	652	16	O8L951	Q8l951 arabidopsis
580	6	0.8	104	16	O8CJ94	Q8cyj94 streptococc	653	16	O8L951	Q8l951 arabidopsis
581	6	0.8	104	16	O8CN70	Q8cn70 staphylococ	654	16	O8L951	Q8l951 arabidopsis
582	6	0.8	105	2	Q8KSF0	Q8ksf0 nostoc sp.	655	16	O8L951	Q8l951 arabidopsis
583	6	0.8	105	16	O8PEC4	Q8pec4 xanthomonas	656	16	O8L951	Q8l951 arabidopsis
584	6	0.8	105	16	O8NX52	Q8nx52 staphylococ	657	16	O8L951	Q8l951 arabidopsis
585	6	0.8	106	2	Q9FCX9	Q9fcx9 bdellovibri	658	16	O8L951	Q8l951 arabidopsis
586	6	0.8	106	9	Q8H911	Q8h911 enterobacte	659	16	O8L951	Q8l951 arabidopsis
587	6	0.8	106	11	O8C8A4	Q8c8a4 mus musculu	660	16	O8L951	Q8l951 arabidopsis
588	6	0.8	107	6	Q9BG41	Q9bg41 sorex arane	661	16	O8L951	Q8l951 arabidopsis
589	6	0.8	107	17	O58937	O58937 pyrococcus	662	16	O8L951	Q8l951 arabidopsis
590	6	0.8	108	16	O8F9J4	Q8f9j4 leptospira	663	16	O8L951	Q8l951 arabidopsis
591	6	0.8	108	16	O8EYX7	Q8eyx7 leptospira	664	16	O8L951	Q8l951 arabidopsis
592	6	0.8	108	17	O59020	O59020 pyrococcus	665	16	O8L951	Q8l951 arabidopsis
593	6	0.8	108	17	Q96YF3	Q96yf3 sulfolobus	666	16	O8L951	Q8l951 arabidopsis
594	6	0.8	109	10	O8L8V1	Q8l8v1 arabidopsis	667	16	O8L951	Q8l951 arabidopsis
595	6	0.8	109	10	O8L8V1	Q8l8v1 arabidopsis	668	16	O8L951	Q8l951 arabidopsis
596	6	0.8	109	10	O8L8V1	Q8l8v1 arabidopsis	669	16	O8L951	Q8l951 arabidopsis
597	6	0.8	109	16	O8L8V1	Q8l8v1 arabidopsis	670	16	O8L951	Q8l951 arabidopsis
598	6	0.8	110	2	Q9VQ33	Q9vq33 escherichia	671	16	O8L951	Q8l951 arabidopsis
599	6	0.8	110	2	Q8VNS4	Q8vns4 escherichia	672	16	O8L951	Q8l951 arabidopsis
600	6	0.8	110	5	Q9NJT0	Q9njt0 trypanosoma	673	16	O8L951	Q8l951 arabidopsis

674	129	3	Q66720	Q66720 saccharomyc	747	6	0.8	142	16	Q8SZT9	Q8etz9 leptospira
675	129	6	P79326	P79326 sus scrofa	748	6	0.8	143	9	Q99365	Q99365 bacterioph
676	129	10	Q8W1Q7	Q8W1Q7 oryza nivar	749	6	0.8	143	16	Q8EQ52	Q8eq52 oceanobacil
677	129	10	Q8H8L8	Q8H8L8 oryza sativ	750	6	0.8	144	11	Q922T3	Q922t3 mus musculu
678	129	16	Q9RSX9	Q9RSX9 dieinococcu	751	6	0.8	144	11	Q93UW2	Q93uw2 magnetospir
679	129	16	Q92PU6	Q92pu6 rhizobium m	752	6	0.8	145	4	Q8N888	Q8n888 homo sapien
680	129	16	Q8d9R9	Q8d9r9 vibrio vuln	753	6	0.8	145	11	Q8BNX9	Q8bnx9 mus musculu
681	130	16	Q8YXP4	Q8ypx4 anabaena sp	754	6	0.8	145	13	Q9DFE4	Q9dfE4 oncorhynch
682	130	16	Q50393	Q50393 mycobacteri	755	6	0.8	145	16	Q9KCU1	Q9kcU1 bacillus ha
683	130	16	Q99W96	Q99w96 staphylococ	756	6	0.8	145	17	Q9YE00	Q9ye00 aeropyrum p
684	131	2	Q93Q25	Q93q25 clostridium	757	6	0.8	146	2	Q33514	Q33514 rhodobacter
685	131	10	Q8LC28	Q8lc28 arabidopsis	758	6	0.8	146	4	Q9NVD2	Q9nvD2 homo sapien
686	131	11	Q8C5T1	Q8c5t1 mus musculu	759	6	0.8	146	4	Q9GZ44	Q9gz44 homo sapien
687	131	16	Q9KC67	Q9kc67 bacillus ha	760	6	0.8	146	9	Q9HAB7	Q9hab7 homo sapien
688	131	16	Q9HXA9	Q9hxa9 pseudomonas	761	6	0.8	146	9	Q37836	Q37836 bacterioph
689	131	16	Q8PFF9	Q8pf59 xanthomonas	762	6	0.8	146	11	Q8CAJ2	Q8caj2 mus musculu
690	132	2	Q9AQN1	Q9aqn1 pseudomonas	763	6	0.8	146	16	Q8Z341	Q8z341 salmonella
691	132	5	Q9BJC8	Q9bjc8 leishmania	764	6	0.8	146	16	Q92MG6	Q92mg6 rhizobium m
692	132	17	Q9YCE2	Q9yc62 aeropyrum p	765	6	0.8	146	17	Q8ULU8	Q8ulu8 pyrococcus
693	133	2	Q8XPP6	Q8kpp6 synechococc	766	6	0.8	147	4	Q8N5U1	Q8n5u1 homo sapien
694	133	2	Q8VPE6	Q8vpv6 synechococc	767	6	0.8	147	5	Q9VE61	Q9ve61 drosophila
695	133	11	Q8BPB4	Q8bpb4 mus musculu	768	6	0.8	147	5	Q61641	Q61641 onchocerca
696	133	12	Q8VB52	Q8vb52 white spot	769	6	0.8	147	8	Q94XV7	Q94xv7 sistrurus c
697	133	16	Q8HTK6	Q8htk6 pseudomonas	770	6	0.8	147	8	Q94XV9	Q94xv9 agkistrodon
698	134	2	Q8CQE6	Q8cqe6 rhizobium t	771	6	0.8	147	8	Q94XV8	Q94xv8 agkistrodon
699	134	4	Q8NAD3	Q8nad3 homo sapien	772	6	0.8	147	16	P96272	P96272 mycobacteri
700	134	10	Q9C944	Q9c944 arabidopsis	773	6	0.8	147	16	Q9RZL9	Q9rzl9 deinoococcus
701	134	16	Q9KLA8	Q9kla8 vibrio chol	774	6	0.8	147	16	Q9RZL9	Q9rzl9 deinoococcus
702	134	16	Q93OC2	Q93oc2 rhizobium m	775	6	0.8	147	16	Q9F3D6	Q9f3d6 streptomyce
703	134	16	Q9YKA2	Q9yka2 anabaena sp	776	6	0.8	147	16	Q8DMF3	Q8dmf3 synechococc
704	135	5	Q8MKZ6	Q8mkz6 drosophila	777	6	0.8	148	2	Q50574	Q50574 bacillus ps
705	135	10	Q8LT04	Q8lt04 oryza sativ	778	6	0.8	148	4	Q8N7B4	Q8n7b4 homo sapien
706	135	10	Q8S5S1	Q8s5s1 oryza sativ	779	6	0.8	148	4	Q8ND38	Q8nd38 homo sapien
707	135	16	Q9RJ72	Q9rj72 streptomyce	780	6	0.8	148	13	Q8AY66	Q8ay66 gallus gall
708	135	17	Q8TK43	Q8tk43 methanosarc	781	6	0.8	148	16	Q92WI2	Q92wi2 rhizobium m
709	136	2	Q8G126	Q8gi26 pseudomonas	782	6	0.8	148	16	Q92UB4	Q92ub4 rhizobium m
710	136	10	Q23628	Q23628 arabidopsis	783	6	0.8	149	4	Q9BZ31	Q9bz31 homo sapien
711	136	10	Q9AYK6	Q9ayk6 oryza sativ	784	6	0.8	149	12	Q9LIH8	Q9lin8 frog adenov
712	136	12	Q8QHP4	Q8qhp4 human cytom	785	6	0.8	149	16	Q9KRM6	Q9krm6 vibrio chol
713	136	12	Q9E7L7	Q9e7l7 human cytom	786	6	0.8	149	16	Q9K9X6	Q9k9x6 bacillus h
714	136	12	Q9E7L8	Q9e7l8 human cytom	787	6	0.8	149	16	Q9L1M5	Q9llm5 streptomyce
715	136	16	P74444	P74444 synechocyst	788	6	0.8	149	16	Q8FCX1	Q8fcx1 corynebacte
716	136	16	Q9WZAL	Q9wzal thermotoga	789	6	0.8	150	4	Q96AJ4	Q96aj4 homo sapien
717	136	16	Q8EXL6	Q8exl6 leptospira	790	6	0.8	150	4	Q9H5Q1	Q9h5q1 homo sapien
718	137	4	Q9UDR1	Q9udr1 homo sapien	791	6	0.8	150	16	Q97HB2	Q97hb2 clostridium
719	138	1	Q9HH46	Q9hh46 methanobact	792	6	0.8	150	16	Q92GW1	Q92gw1 rickettsia
720	138	10	Q94CC9	Q9acc9 arabidopsis	793	6	0.8	150	16	Q8XO07	Q8xo07 ralstonia s
721	138	10	Q8H7Y8	Q8h7y8 oryza sativ	794	6	0.8	150	16	Q8X7M1	Q8x7m1 escherichia
722	138	11	Q70329	Q70329 mesocricetu	795	6	0.8	150	16	Q8FCQ6	Q8fcq6 corynebacte
723	138	16	Q9RD30	Q9rd30 streptomyce	796	6	0.8	150	16	Q8DTS6	Q8dts6 streptococc
724	139	4	Q96QN2	Q96qn2 homo sapien	797	6	0.8	150	16	Q8CR96	Q8cr96 staphylococ
725	139	5	Q8T1X1	Q8tlx1 dictyosteli	798	6	0.8	151	4	Q9UF74	Q9uf74 homo sapien
726	139	8	Q94XV6	Q94xv6 sistrurus m	799	6	0.8	151	5	Q77243	Q77243 dictyosteli
727	139	10	Q9S857	Q9s857 oryza sativ	800	6	0.8	151	17	Q8PXS6	Q8pxs6 methanosarc
728	139	10	Q9SNZ3	Q9snz3 oryza sativ	801	6	0.8	152	2	Q8RRH9	Q8rrh9 helicobacte
729	139	16	Q91524	Q91524 pseudomonas	802	6	0.8	152	2	Q8GNY6	Q8gny6 staphylococ
730	139	16	Q97FG9	Q97fg9 clostridium	803	6	0.8	152	3	Q96U94	Q96u94 neurospora
731	139	16	Q8XKD4	Q8xkd4 clostridium	804	6	0.8	152	4	Q9UIG5	Q9uig5 homo sapien
732	139	17	Q29174	Q29174 archaeoglob	805	6	0.8	152	16	Q9JQ02	Q9jq02 neisseria m
733	139	17	Q97YB6	Q97yb6 sulfolobus	806	6	0.8	152	16	Q8YJC8	Q8yjc8 brucella me
734	139	17	Q8ZXJ2	Q8zxj2 pyrobaculum	807	6	0.8	152	16	Q8FYU8	Q8fyu8 brucella su
735	140	10	Q49628	Q49628 arabidopsis	808	6	0.8	152	16	Q8FVH0	Q8fvyh0 brucella su
736	140	16	Q9RKW0	Q9rkwo streptomyce	809	6	0.8	152	17	Q9YEH2	Q9yeh2 aeropyrum p
737	140	16	Q8FRC1	Q8frc1 corynebacte	810	6	0.8	153	2	Q8VR74	Q8vr74 escherichia
738	141	2	Q87687	Q87687 blastobacte	811	6	0.8	153	5	Q9NA70	Q9na70 caenorhabdi
739	141	2	Q66334	Q66334 unidentified	812	6	0.8	153	8	Q9B8U2	Q9b8u2 schistosoma
740	141	5	Q95KD3	Q95kd3 drosophila	813	6	0.8	153	16	Q92S80	Q92s80 rhizobium m
741	141	6	Q95KD3	Q95kd3 macaca fasc	814	6	0.8	153	16	Q8VSA6	Q8vsa6 anabaena sp
742	141	11	Q9CW33	Q9cw33 mus musculu	815	6	0.8	153	16	Q8YBP0	Q8ypb0 anabaena sp
743	141	16	Q8YJAO	Q8yja0 brucella me	816	6	0.8	153	16	Q8RFK6	Q8rfk6 fusobacteri
744	142	2	Q93PL3	Q93pl3 pseudomonas	817	6	0.8	153	16	Q8DMT6	Q8dmt6 synechococc
745	142	3	Q05169	Q05169 saccharomyc	818	6	0.8	154	3	Q3C0S4	Q3c0s4 cryptococc
746	142	16	Q92LL6	Q92ll6 rhizobium m	819	6	0.8	154	4	Q8NAT4	Q8nat4 homo sapien
								154	5	Q8AIQ6	Q8aiqu6 drosophila

820	6	0.8	154	9	Q8W628	Q8W628 cyanophage	893	6	0.8	168	13	Q9PSG9	Q9PSG9 xenopus lae
821	6	0.8	154	16	Q8XUB9	Q8XUB9 ralsstonia s	894	6	0.8	168	16	Q92KW7	Q92KW7 rhizobium m
822	6	0.8	154	16	Q8X9U6	Q8X9U6 escherichia	895	6	0.8	169	5	Q92KW7	Q92KW7 rhizobium m
823	6	0.8	154	16	Q8RFX5	Q8RFX5 fusobacteri	896	6	0.8	169	5	Q92KW7	Q92KW7 rhizobium m
824	6	0.8	154	17	Q8ZSS7	Q8ZSS7 pyrobaculum	897	6	0.8	169	8	Q8MX22	Q8MX22 agomon exte
825	6	0.8	155	5	Q02421	Q02421 bdellooura c	898	6	0.8	169	11	Q9CYM0	Q9CYM0 mus musculu
826	6	0.8	155	5	Q95ZQ7	Q95ZQ7 caenorhabdi	899	6	0.8	169	11	Q9CYM0	Q9CYM0 mus musculu
827	6	0.8	155	12	Q65720	Q65720 barley stri	900	6	0.8	169	11	Q9DCR4	Q9DCR4 mus musculu
828	6	0.8	155	16	Q8T64	Q8T64 vibrio chol	901	6	0.8	169	11	Q64364	Q64364 mus musculu
829	6	0.8	155	16	Q8Z5R6	Q8Z5R6 salmonella	902	6	0.8	169	12	Q91I79	Q91I79 hepatitis c
830	6	0.8	155	16	Q8FJY3	Q8FJY3 escherichia	903	6	0.8	169	12	Q917S6	Q917S6 hepatitis c
831	6	0.8	156	16	Q8ZNT1	Q8ZNT1 salmonella	904	6	0.8	169	12	Q917S5	Q917S5 hepatitis c
832	6	0.8	156	16	Q8Z507	Q8Z507 salmonella	905	6	0.8	169	12	Q91H28	Q91H28 influenza a
833	6	0.8	156	16	Q8DR29	Q8DR29 streptococc	906	6	0.8	169	12	Q917S7	Q917S7 hepatitis c
834	6	0.8	157	2	Q93K50	Q93K50 klebsiella	907	6	0.8	169	12	Q91H25	Q91H25 influenza a
835	6	0.8	157	11	Q8CLM1	Q8CLM1 mus musculu	908	6	0.8	169	12	Q91H29	Q91H29 influenza a
836	6	0.8	157	11	Q9WZM7	Q9WZM7 thermotoga	909	6	0.8	169	12	Q91H27	Q91H27 influenza a
837	6	0.8	157	16	Q8XVZ6	Q8XVZ6 ralsstonia s	910	6	0.8	169	12	Q91H26	Q91H26 influenza a
838	6	0.8	158	2	Q8VNN8	Q8VNN8 escherichia	911	6	0.8	169	12	Q91H24	Q91H24 influenza a
839	6	0.8	158	10	Q8H1A6	Q8H1A6 pisum sativ	912	6	0.8	170	2	Q8GNV7	Q8GNV7 staphylococ
840	6	0.8	158	11	Q8BLJ9	Q8BLJ9 mus musculu	913	6	0.8	170	5	Q8WYU0	Q8WYU0 morion arid
841	6	0.8	158	16	Q9KDI9	Q9KDI9 bacillus ha	914	6	0.8	170	13	Q8AXW4	Q8AXW4 anguilla an
842	6	0.8	158	16	Q9CCM4	Q9CCM4 mycobacteri	915	6	0.8	171	2	Q44242	Q44242 anabaena sp
843	6	0.8	158	16	Q9CBH3	Q9CBH3 rhizobium l	916	6	0.8	171	5	Q8ITE9	Q8ITE9 choristoneu
844	6	0.8	158	16	Q93Q48	Q93Q48 clostridium	917	6	0.8	171	11	Q9D389	Q9D389 mus musculu
845	6	0.8	159	11	Q9CQ17	Q9CQ17 mus musculu	918	6	0.8	171	16	P74151	P74151 synechocyst
846	6	0.8	159	11	Q8BP25	Q8BP25 mus musculu	919	6	0.8	171	16	Q8Z120	Q8Z120 salmonella
847	6	0.8	160	4	Q6AR6	Q6AR6 homo sapien	920	6	0.8	171	16	Q8EJD4	Q8EJD4 shewanella
848	6	0.8	160	10	Q8W259	Q8W259 oryza sativ	921	6	0.8	172	5	Q23248	Q23248 caenorhabdi
849	6	0.8	160	10	Q949B5	Q949B5 oryza sativ	922	6	0.8	172	5	Q25313	Q25313 locusta mig
850	6	0.8	160	16	Q9JQY2	Q9JQY2 neisseria m	923	6	0.8	172	5	Q8IGZ5	Q8IGZ5 drosophila
851	6	0.8	160	16	Q8FG23	Q8FG23 escherichia	924	6	0.8	172	11	Q9D0E6	Q9D0E6 mus musculu
852	6	0.8	160	16	Q8DIK0	Q8DIK0 synechococc	925	6	0.8	172	16	Q8R867	Q8R867 thermoanaer
853	6	0.8	161	4	Q7TE17	Q7TE17 homo sapien	926	6	0.8	173	8	Q8HLE6	Q8HLE6 corynebacte
854	6	0.8	161	10	Q42785	Q42785 glycine max	927	6	0.8	173	8	Q8HCV4	Q8HCV4 fugu rubrip
855	6	0.8	161	10	Q9FFX4	Q9FFX4 arabidopsis	928	6	0.8	173	12	Q8B261	Q8B261 influenza a
856	6	0.8	161	12	Q8UWN1	Q8UWN1 influenza a	929	6	0.8	173	12	Q8B259	Q8B259 influenza a
857	6	0.8	161	12	Q8JVK4	Q8JVK4 paprika mil	930	6	0.8	173	12	Q8B257	Q8B257 influenza a
858	6	0.8	161	16	Q8XWC5	Q8XWC5 ralsstonia s	931	6	0.8	173	12	Q8B255	Q8B255 influenza a
859	6	0.8	161	16	Q8F354	Q8F354 leptospira	932	6	0.8	173	12	Q8B253	Q8B253 influenza a
860	6	0.8	162	4	Q9UQR7	Q9UQR7 homo sapien	933	6	0.8	173	12	Q8B249	Q8B249 influenza a
861	6	0.8	162	5	Q8SSW2	Q8SSW2 dictyosteli	934	6	0.8	173	12	Q8B247	Q8B247 influenza a
862	6	0.8	162	6	Q9NN3	Q9NN3 papio hamad	935	6	0.8	173	12	Q8B245	Q8B245 influenza a
863	6	0.8	162	9	Q9ZXF0	Q9ZXF0 bacteriopho	936	6	0.8	173	12	Q8B243	Q8B243 influenza a
864	6	0.8	162	16	Q9HTP6	Q9HTP6 pseudomonas	937	6	0.8	173	12	Q8B241	Q8B241 influenza a
865	6	0.8	162	16	Q88FA2	Q88FA2 rhizobium l	938	6	0.8	173	12	Q8B239	Q8B239 influenza a
866	6	0.8	162	16	Q986W8	Q986W8 thizobium l	939	6	0.8	173	12	Q8B237	Q8B237 influenza a
867	6	0.8	162	16	Q51330	Q51330 borrelia bu	940	6	0.8	173	12	Q8B233	Q8B233 influenza a
868	6	0.8	162	16	Q8EMU7	Q8EMU7 oceanobacil	941	6	0.8	173	12	Q8B231	Q8B231 influenza a
869	6	0.8	162	16	Q8DPF6	Q8DPF6 streptococc	942	6	0.8	173	12	Q8B229	Q8B229 influenza a
870	6	0.8	163	10	Q93VP4	Q93VP4 arabidopsis	943	6	0.8	173	12	Q8B227	Q8B227 influenza a
871	6	0.8	163	10	Q9STJ7	Q9STJ7 arabidopsis	944	6	0.8	173	12	Q8B225	Q8B225 influenza a
872	6	0.8	163	11	Q9QWK7	Q9QWK7 rattus norv	945	6	0.8	173	12	Q8B223	Q8B223 influenza a
873	6	0.8	163	12	Q70873	Q70873 equine arte	946	6	0.8	173	12	Q8B221	Q8B221 influenza a
874	6	0.8	163	12	Q53240	Q53240 mycobacteri	947	6	0.8	173	12	Q8B219	Q8B219 influenza a
875	6	0.8	163	16	Q8E300	Q8E300 streptococc	948	6	0.8	173	12	Q8B217	Q8B217 influenza a
876	6	0.8	163	16	Q8DX77	Q8DX77 streptococc	949	6	0.8	173	12	Q8B213	Q8B213 influenza a
877	6	0.8	164	15	Q9PD17	Q9PD17 xylella fas	950	6	0.8	173	12	Q8B211	Q8B211 influenza a
878	6	0.8	164	16	Q99X02	Q99X02 staphylococ	951	6	0.8	173	12	Q8B209	Q8B209 influenza a
879	6	0.8	164	16	Q8FZJ2	Q8FZJ2 brucella su	952	6	0.8	173	12	Q8B207	Q8B207 influenza a
880	6	0.8	164	16	Q8DHV3	Q8DHV3 synechococc	953	6	0.8	173	12	Q8B205	Q8B205 influenza a
881	6	0.8	165	2	Q9X911	Q9X911 versinia en	954	6	0.8	173	12	Q8B201	Q8B201 influenza a
882	6	0.8	165	5	Q18505	Q18505 caenorhabdi	955	6	0.8	173	12	Q8B129	Q8B129 influenza a
883	6	0.8	165	16	Q8Y2Q5	Q8Y2Q5 ralsstonia s	956	6	0.8	173	12	Q8B127	Q8B127 influenza a
884	6	0.8	165	16	Q8NX50	Q8NX50 staphylococ	957	6	0.8	173	12	Q8B125	Q8B125 influenza a
885	6	0.8	166	13	Q98SF1	Q98SF1 xenopus lae	958	6	0.8	173	12	Q8B119	Q8B119 influenza a
886	6	0.8	166	16	Q9RVN0	Q9RVN0 deinococcus	959	6	0.8	173	12	Q8B117	Q8B117 influenza a
887	6	0.8	167	2	Q53060	Q53060 rhodobacter	960	6	0.8	173	12	Q8B115	Q8B115 influenza a
888	6	0.8	167	16	Q9X8H9	Q9X8H9 streptomyce	961	6	0.8	173	12	Q8B113	Q8B113 influenza a
889	6	0.8	168	2	Q32877	Q32877 mycobacteri	962	6	0.8	173	12	Q8B111	Q8B111 influenza a
890	6	0.8	168	2	Q49813	Q49813 mycobacteri	963	6	0.8	173	12	Q8B109	Q8B109 influenza a
891	6	0.8	168	8	Q48255	Q48255 demmyus vau	964	6	0.8	173	12	Q8B107	Q8B107 influenza a
892	6	0.8	168	8	Q9MGA4	Q9MGA4 chrysodidym	965	6	0.8	173	12	Q8B105	Q8B105 influenza a

QY 593 VRYNSGADFYGAEGEIIYFKPTPRYIGVSGDYVRGLKNLPSLPGRDAYGNRP 647
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 599 VRYNSGADFYGAEGEIIYFKPTPRYIGVSGDYVRGLKNLPSLPGRDAYGNRP 653

RESULT 3

Q9CLW8 PRELIMINARY; PRT; 809 AA.
 AC Q9CLW8; (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein PM1081.
 GN PM1081.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70; PubMed=11248100;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
 DR EMBL: AE006149; AAK03165.1; -;
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Rcg; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 809 AA; 90903 MW; 23026B7DFCBF14DB CRC64;

Query Match 1.6%; Score 12; DB 16; Length 809;
 Best Local Similarity 100.0%; Pred. No. 0.0043;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GGGASAPVIRGQ 100
 ||||||||||||
 Db 103 GGGASAPVIRGQ 114

RESULT 4

Q9Y911 PRELIMINARY; PRT; 205 AA.
 AC Q9Y911; (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein APE2473.
 GN APE2473.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kavarabavasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101 (1999).
 DR EMBL: AP000064; BAA81489.1; -;
 DR InterPro: IPR004843; M-ppetrase.
 DR Pfam: PF00149; Metallophos; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 205 AA; 23063 MW; 7A1D29E93615304F CRC64;

Query Match 1.2%; Score 9; DB 17; Length 205;

Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 318 HAHAHNGKP 326
 ||||||||
 Db 162 HAHAHNGKP 170

RESULT 5

Q9AER9 PRELIMINARY; PRT; 766 AA.
 AC Q9AER9; (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Iron-regulated outer membrane protein.
 GN IRP.
 OS Pasteurella haemolytica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Mannheimia.
 OX NCBI_TaxID=75985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lo R.Y., Graham M.R.;
 RT "Putative TonB dependent receptor of Mannheimia haemolytica.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL: AY028475; AAK39743.1; -;
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Rcg; 1.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
 DR Membrane; Outer membrane; Receptor; TonB box.
 KW SEQUENCE 766 AA; 86440 MW; 7D76FA05D7C5619 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 766;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 QELYAHGKH 523
 ||||||||
 Db 521 QELYAHGKH 529

RESULT 6

Q9ZH42 PRELIMINARY; PRT; 93 AA.
 AC Q9ZH42; (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE Stability protein StbE.
 GN STBE.
 OS Morganella morganii (Proteus morganii).
 OG Plasmid R485.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Morganella.
 OX NCBI_TaxID=582;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99047595; PubMed=9829958;
 RA Hayes F.;
 RT "A family of stability determinants in pathogenic bacteria.";
 RL J. Bacteriol. 180:6415-6418 (1998).
 DR EMBL: AF072126; AAC84004.1; -;
 KW Plasmid.
 SQ SEQUENCE 93 AA; 11024 MW; 373AB3217769BDF2 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 93;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 PRVPAARL 664
 ||||||||


```

Db          37 PRVPAARL 44

RESULT 7
Q92EP9      PRELIMINARY;      PRT;      99 AA.

ID Q92EP9;
AC Q92EP9;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein lin0409.
GN LIN0409.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95642.1; -.
DR Listlist; LIN00409; -.
SQ SEQUENCE 99 AA; 11124 MW; 894D5703DCFE65B1 CRC64;

Query Match      1.1%; Score 8; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 11;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      2 AQTILKPI 9
Db          46 AQTILKPI 53

RESULT 8
Q8Y9X6      PRELIMINARY;      PRT;      99 AA.

ID Q8Y9X6;
AC Q8Y9X6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein lmo0391.
GN LMO0391.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95642.1; -.
DR Listlist; LIN00409; -.
SQ SEQUENCE 99 AA; 11124 MW; 894D5703DCFE65B1 CRC64;

Query Match      1.1%; Score 8; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 11;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      2 AQTILKPI 9
Db          46 AQTILKPI 53

RESULT 8
Q8Y9X6      PRELIMINARY;      PRT;      99 AA.

ID Q8Y9X6;
AC Q8Y9X6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein lmo0391.
GN LMO0391.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95642.1; -.
DR Listlist; LIN00409; -.
SQ SEQUENCE 99 AA; 11124 MW; 894D5703DCFE65B1 CRC64;

Query Match      1.1%; Score 8; DB 11; Length 171;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      479 AHRQTARS 486
Db          93 AHRQTARS 100

RESULT 10
Q9VNN6      PRELIMINARY;      PRT;      268 AA.

ID Q9VNN6;
AC Q9VNN6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

```

```

DE  CG1155 protein (LD24139P).
GN  SP558 OR CG1155.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celisniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galie R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush P., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimball B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Peinert K., Rittington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster."
RL  Science 287:2185-2195(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  STRAIN=Berkeley;
RC  Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA  Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA  Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA  Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA  Yu C., Lewis S.E., Rubin G.M., Celisniker S.;
RL  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AE003600; AAF51893.1; -
DR  EMBL; AY058538; AAU13767.1; -
DR  FlyBase; FBgn0040279; SP558.
SQ  SEQUENCE 268 AA; 2798 MW; 2A3B561F3B316B3 CRC64;

Query Match 1.1%; Score 8; DB 5; Length 268;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 RSSNIELA 547
Db 67 RSSNIELA 74

RESULT 11

Query Match 1.1%; Score 8; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 13
Q93UWI

```

```
ID Q93UW1 PRELIMINARY; PRT; 399 AA.
AC Q93UW1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NAD+ dependent formate dehydrogenase.
GN FDH.
OS Hyphomicrobium sp. JCI17.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiaceae; Hyphomicrobium.
OX NCBI_TaxID=142666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCI17;
RA Tanaka Y., Yoshida T., Watanabe K., Mitsunaga T.;
RT "Characterization, gene cloning and expression of NAD+-dependent
RT formate dehydrogenase from a methylophilic bacterium Hyphomicrobium
RT sp. JCI17."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051073; BAB55449.1; -.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 399 AA; 44044 MW; 853BAD2C635864DC CRC64;

Query Match 1.1%; Score 8; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 54 VSGELGLR 61

RESULT 14
Q17898 PRELIMINARY; PRT; 399 AA.
ID Q17898;
AC Q17898;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C10C5.3 protein.
GN C10C5.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018 (1998).
DR EMBL; Z68214; CAA92445.1; -.
DR MEROPS; M20.973; -.
DR WormPep; C10C5.3; CE03011.
DR InterPro; IPR001261; ARGE_DAPE_CPG2.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; 1.
DR PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
SQ SEQUENCE 399 AA; 45679 MW; 8036854C8D0A6D8B CRC64;

Query Match 1.1%; Score 8; DB 5; Length 399;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 15 LINTPLLA 22
Db 361 LINTPLLA 368

RESULT 15
Q9RSV0 PRELIMINARY; PRT; 400 AA.
ID Q9RSV0;
AC Q9RSV0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NAD(+) dependent formate dehydrogenase (EC 1.2.1.2) (Fragment).
OS Mycobacterium vaccae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1810;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96158475; PubMed=8597552;
RA Galkin A., Kulakova L., Tishkov V., Esaki N., Soda K.;
RT "Cloning of formate dehydrogenase gene from a methanol-utilizing
RT bacterium Mycobacterium vaccae N10."
RL Appl. Microbiol. Biotechnol. 44:479-483 (1995).
DR HSSP; P33160; 2NAD.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 400 AA; 44018 MW; 44228E88CA22BD9B CRC64;

Query Match 1.1%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 16
Q93GW3 PRELIMINARY; PRT; 400 AA.
ID Q93GW3;
AC Q93GW3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NAD-dependent formate dehydrogenase.
GN FDH.
OS Paracoccus sp. 12-A.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxID=171192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12-A;
RA Shinoda T., Satoh T., Mineki S., Iida M., Taguchi H.;
RT "Cloning, nucleotide sequence and expression in Escherichia coli of
RT the gene for formate dehydrogenase of Paracoccus sp. 12-A, a formate-
RT assimilating bacterium."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071373; BAB64941.1; -.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 400 AA; 43757 MW; 13195AED9AC299D8 CRC64;
```

Query Match		1.1%; Score 8; DB 2; Length 400;
Best Local Similarity		100.0%; Pred. No. 39;
Matches		8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	172 VSGELGLR 179	
DB	54 VSGELGLR 61	
RESULT 17		
Q93GV1	PRELIMINARY;	PRT; 401 AA.
AC	Q93GV1;	
DT	01-DEC-2001 (TRENBLrel. 19, Created)	
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)	
DT	01-MAR-2003 (TRENBLrel. 23, Last annotation update)	
DE	Formate dehydrogenase (EC 1.2.1.2).	
GN	FDH.	
OS	Mycobacterium vaccae.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Cornebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1810;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Galkin A., Kulakova L., Tishkov V., Esaki N., Soda K.;	
RA	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	MEDLINE=96158475; PubMed=8597552;	
RT	"Cloning of formate dehydrogenase gene from a methanol-utilizing	
RT	Bacterium Mycobacterium vaccae N10.";	
RT	Appl. Microbiol. Biotechnol. 44:479-483 (1995).	
DE	EMBL; AB072394; BAB69476.1; -.	
DR	InterPro; IPR006139; 2-Hacid_DH.	
DR	InterPro; IPR006140; 2-Hacid_DH_C.	
DR	Pfam; PF00389; 2-Hacid_DH; 1.	
DR	Pfam; PF02826; 2-Hacid_DH_C; 1.	
DR	PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.	
DR	PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.	
DR	PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.	
KW	Oxidoreductase.	
SQ	SEQUENCE 401 AA; 44149 MW; 8C4821B4790AAD26 CRC64;	
Query Match		
Best Local Similarity		
Matches		
QY	172 VSGELGLR 179	
DB	54 VSGELGLR 61	
RESULT 18		
Q930E7	PRELIMINARY;	PRT; 401 AA.
AC	Q930E7;	
DT	01-DEC-2001 (TRENBLrel. 19, Created)	
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)	
DT	01-MAR-2003 (TRENBLrel. 23, Last annotation update)	
DE	Probable NAD-dependent formate dehydrogenase.	
GN	RA0251 OR SWA0478.	
OS	Rhizobium meliloti (sinorhizobium meliloti).	
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;	
OC	Rhizobiaceae; Sinorhizobium.	
OX	NCBI_TaxID=382;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=1021;	
RX	MEDLINE=21396509; PubMed=11481432;	
RA	Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,	
Query Match		1.1%; Score 8; DB 2; Length 401;
Best Local Similarity		100.0%; Pred. No. 39;
Matches		8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	172 VSGELGLR 179	
DB	54 VSGELGLR 61	
RESULT 19		
Q98375	PRELIMINARY;	PRT; 402 AA.
AC	Q98375;	
DT	01-JUL-1997 (TRENBLrel. 04, Created)	
DT	01-JUL-1997 (TRENBLrel. 04, Last sequence update)	
DT	01-MAR-2003 (TRENBLrel. 23, Last annotation update)	
DE	NAD-dependent formate dehydrogenase (EC 1.2.1.2).	
GN	FDH.	
OS	Moraxella sp.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
OC	Moraxellaceae; Moraxella.	
OX	NCBI_TaxID=479;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-1;	
RA	Galkin A.G.;	
RA	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; Y13245; CAA73696.1; -.	
DR	HSP; E33160; 2NAD.	
DR	InterPro; IPR006139; 2-Hacid_DH.	
DR	InterPro; IPR006140; 2-Hacid_DH_C.	
DR	Pfam; PF00389; 2-Hacid_DH; 1.	
DR	Pfam; PF02826; 2-Hacid_DH_C; 1.	
DR	PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.	
DR	PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.	
DR	PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.	
KW	NAD; Oxidoreductase.	
SQ	SEQUENCE 402 AA; 44143 MW; C511CE1CE996BBB CRC64;	
Query Match		
Best Local Similarity		
Matches		
QY	172 VSGELGLR 179	
DB	54 VSGELGLR 61	
RESULT 20		
Q9F771	PRELIMINARY;	PRT; 427 AA.
ID	Q9F771;	
AC	Q9F771;	
DT	01-MAR-2001 (TRENBLrel. 16, Created)	
DT	01-MAR-2001 (TRENBLrel. 16, Last sequence update)	
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)	

DE Xerc.
 GN XERC.
 OS Pseudomonas aeruginosa.
 OC Plasmid pKLC102.
 DT Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 DE Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=SG17M;
 RX MEDLINE=20476411; PubMed=11021913;
 RA Kiewitz C., Larbig K., Klockgether J., Weinel C., Tummler B.;
 RT "Monitoring genome evolution ex vivo: reversible chromosomal
 RT segregation of a 106 kb plasmid at two trna(Lys) gene loci in
 RT sequential Pseudomonas aeruginosa airway isolates.";
 RL Microbiology 146:2365-2373(2000).
 DR EMBL; AF285416; AA02084.1; -;
 DR InterPro; IPR002104; Phage_integrase.
 DR Pfam; PF00589; Phage_integrase; 1.
 KW Plasmid.
 SQ SEQUENCE 427 AA; 49137 MW; 830637E70A0420A7 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 427;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 HRAVLQWR 245
 |||||
 Db 46 HRAVLQWR 53

RESULT 21

ID P74655 PRELIMINARY; PRT; 544 AA.
 AC P74655;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein sll1550.
 GN SLL1550.
 OS Synecocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90917; BAA18771.1; -;
 DR InterPro; IPR001119; SLH.
 DR Pfam; PF00395; SLH; 1.
 DR PROSITE; PS01072; SLH DOMAIN; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 544 AA; 58631 MW; 48A4905CA184F61A CRC64;

Query Match 1.1%; Score 8; DB 16; Length 544;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 RLSSGNLE 186
 |||||
 Db 222 RLSSGNLE 229

RESULT 22

Q9CMR7

ID Q9CMR7 PRELIMINARY; PRT; 925 AA.
 AC Q9CMR7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein PM0745.
 GN PM0745.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL; A5006112; AAK02829.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep Rec; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 925 AA; 105081 MW; F46A9F52B59145E3 CRC64;

Query Match 1.1%; Score 8; DB 16; Length 925;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 GDYVRGRL 630
 |||||
 Db 696 GDYVRGRL 703

RESULT 23

ID 074360 PRELIMINARY; PRT; 933 AA.
 AC 074360;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE GTPase-activator protein for Rho-like GTPases.
 GN SPBC28E12.03.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972h-;
 RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.

EMBL; AL031517; CAA20650.2; -;
 DR GenesDB Spombe; SPEC28E12.03; -;
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR00198; RhoGAP.
 DR Pfam; PF00412; LIM; 2.
 DR Pfam; PF00620; RhoGAP; 1.
 DR ProDom; PD000094; LIM; 1.
 DR SMART; SM00132; LIM; 2.
 DR SMART; SM00324; RhoGAP; 1.
 DR PROSITE; PS00478; LIM DOMAIN 1; 1.
 DR PROSITE; PS50023; LIM DOMAIN 2; 1.
 KW LIM domain; Metal-binding; Zinc.
 SQ SEQUENCE 933 AA; 104487 MW; 2EE75150F62B5586 CRC64;

Query Match 1.1%; Score 8; DB 3; Length 933;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYRKSG 213

```

DB 779 EGLYRKSG 786
|||||
RESULT 24
Q8PD15 PRELIMINARY; PRT; 984 AA.
ID Q8PD15
AC Q8PD15
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein XCC0531.
GN XCC0531.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Perro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spinola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
Tindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
Stubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE012151; AAM39847.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 984 AA; 108243 MW; 07603CAB01E0DC77 CRC64;

Query Match 1.1%; Score 8; DB 16; Length 984;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 747 GRGFTGGV 754
|||||
DB 703 GRGFTGGV 710
|||||
RESULT 25
Q06407 PRELIMINARY; PRT; 1009 AA.
ID Q06407
AC Q06407;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Similar in C-terminus to chimaerin.
GN RGA2 OR D9481.4 OR YDR379W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=S288C;
RA Ding H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[2]

SEQUENCE FROM N.A.
RP STRAIN=S288C;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
Pavello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
Johnston L., Langston Y., Latreille P., Le T., Jordis E., Meneses S.,
Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
Taich A., Trevisan E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
Wilson R., Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RP STRAIN=S288C;
RA Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RP STRAIN=S288C;
RA Jia Y., Cherry J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U28373; AAB64815.1; -.
DR HSSP; P27986; 1PBW.
DR SGD; S0002787; RGA2.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001198; RhoGAP.
DR Pfam; PF00412; LIM; 2.
DR Pfam; PF00620; RhoGAP; 1.
DR ProDom; PD000094; LIM; 2.
DR SMART; SM00132; LIM; 2.
DR SMART; SM00324; RhoGAP; 1.
DR PROSITE; PS00478; LIM DOMAIN 1; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 1009 AA; 113290 MW; FA13BALDF8B35D98 CRC64;

Query Match 1.1%; Score 8; DB 3; Length 1009;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYRKSG 213
|||||
DB 822 EGLYRKSG 829
|||||
RESULT 26
Q8H614 PRELIMINARY; PRT; 2319 AA.
ID Q8H614;
AC Q8H614;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative gag-pol polyprotein.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CV. BSS53;
RX MEDLINE=22253764; PubMed=12368247;
RA Song R., Liaca V., Messing J.;
RT "Mosaic Organization of Orthologous Sequences in Grass Genomes.";
RL Genome Res. 12:1549-1555(2002).
DR EMBL; AF528565; AAN40030.1; -.
KW Polyprotein.
SQ SEQUENCE 2319 AA; 263271 MW; 56378EC6E16EDF08 CRC64;

Query Match 1.1%; Score 8; DB 10; Length 2319;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 ASAPVIRG 99
|||||

```

Db 1807 ASAPVIRG 1814

RESULT 27

Q8GBX6 PRELIMINARY; PRT; 5017 AA.

ID Q8GBX6 AC Q8GBX6

DT 01-MAR-2003 (TEMBLrel. 23, Created)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Polypeptide synthase.

GN SPIH.

OS Polyangium cellulosum.

OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;

OC Sorangineae; Polyangiaceae; Polyangium.

OX NCBI_TaxID=56;

RN [1]

RP SEQUENCE FROM N.A.

RA Knauber J., Bloeker H., Mueller R., Nordieck G., Beyer S.;

RT "The spirangine synthase from Sorangium cellulosum So ce90 - Module

RT duplications as strategy behind the evolution of polyketide

RT synthases.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505006; CAD43449.1; -

SQ SEQUENCE 5017 AA; 525555 MW; 08BED56182669493 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 5017;

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 GLVDVADG 161

Db 2703 GLVDVADG 2710

RESULT 28

Q8KY74 PRELIMINARY; PRT; 48 AA.

ID Q8KY74 AC Q8KY74

DT 01-OCT-2002 (TEMBLrel. 22, Created)

DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Alpha-crystallin protein (Fragment).

GN ACR.

OS Mycobacterium habana.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1784;

RN [1]

RP SEQUENCE FROM N.A.

RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaeels F.;

RT "Detection of a gene encoding an alpha-crystallin homolog in numerous

RT mycobacterial species other than Mycobacterium tuberculosis.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF253458; AAM6920.1; -

DR InterPro; IPR002068; Hsp20.

DR Pfam; PF00011; HSP20; 1.

FT NON TER 1

SQ SEQUENCE 48 AA; 5491 MW; 14DD8B8694296C71 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340

Db 23 RYELRAE 29

RESULT 29

Q8KY83 PRELIMINARY; PRT; 48 AA.

ID Q8KY83 AC Q8KY83

DT 01-OCT-2002 (TEMBLrel. 22, Created)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Alpha-crystallin protein (Fragment).

GN ACR.

OS Mycobacterium triviale.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1798;

RN [1]

RP SEQUENCE FROM N.A.

RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaeels F.;

RT "Detection of a gene encoding an alpha-crystallin homolog in numerous

RT mycobacterial species other than Mycobacterium tuberculosis.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF253451; AAM69213.1; -

DR EMBL; AF253452; AAM69214.1; -

DR InterPro; IPR002068; Hsp20.

DR Pfam; PF00011; HSP20; 1.

FT NON TER 1

SQ SEQUENCE 48 AA; 5477 MW; 7FB68B8694296820 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340

Db 23 RYELRAE 29

RESULT 30

Q8KHQ1 PRELIMINARY; PRT; 48 AA.

ID Q8KHQ1 AC Q8KHQ1

DT 01-OCT-2002 (TEMBLrel. 22, Created)

DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Alpha-crystallin protein (Fragment).

GN ACR.

OS Mycobacterium ulcerans.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1809;

RN [1]

RP SEQUENCE FROM N.A.

RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaeels F.;

RT "Detection of a gene encoding an alpha-crystallin homolog in numerous

RT mycobacterial species other than Mycobacterium tuberculosis.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF253450; AAM69212.1; -

DR EMBL; AF253451; AAM69213.1; -

DR EMBL; AF253452; AAM69214.1; -

DR InterPro; IPR002068; Hsp20.

DR Pfam; PF00011; HSP20; 1.

FT NON TER 1

SQ SEQUENCE 48 AA; 5477 MW; 7FB68B8694296820 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340

Db 23 RYELRAE 29

RESULT 31

Q08482 PRELIMINARY; PRT; 55 AA.

ID Q08482 AC Q08482


```

DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-FEB-1997 (TrEMBLrel. 02, Last annotation update)
DE L protein (Fragment).
GN L.
OS Viral hemorrhagic septicemia virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Novirhabdovirus.
OX NCBI_TaxID=11287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FI13;
RA Schuetze H.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X73873; CAA52082.1; -.
FT NON TER 55
SQ SEQUENCE 55 AA; 6352 MW; 934733D8EEA24A1E CRC64;

Query Match 0.9%; Score 7; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 HOERLPS 513
DB 11 HOERLPS 17

RESULT 32
Q8L411 PRELIMINARY; PRT; 60 AA.
AC Q8L411;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE QJ1103.E04.11 protein (P0675B10.1 protein).
GN QJ1103.E04.11 OR P0675B10.1.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
clone:OJ1103.E04.11";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone:P0675B10.1";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003806; BAC06954.1; -.
DR EMBL; AF004347; BAC07130.1; -.
DR Gramene; Q8L411; -.
SQ SEQUENCE 60 AA; 6630 MW; C799493C3836E719 CRC64;

Query Match 0.9%; Score 7; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 RLPDSR 231
DB 12 RLPDSR 18

RESULT 33
Q9JSV8 PRELIMINARY; PRT; 76 AA.
AC Q9JSV8;

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein NMA2114.
GN NMA2114.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RA MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leathers S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB85328.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 76 AA; 8826 MW; 9DB776C1C687DB48 CRC64;

Query Match 0.9%; Score 7; DB 16; Length 76;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 DRIDANL 680
DB 17 DRIDANL 23

RESULT 34
Q8KY84 PRELIMINARY; PRT; 86 AA.
AC Q8KY84;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium nonchromogenicum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19530;
RA Shampura I.C., Bastian I., Fonteyne P.-A., Portals P.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253445; AAM69207.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON TER 86
FT NON TER 86
SQ SEQUENCE 86 AA; 9722 MW; BAF085B56BF1DBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 35

```

```
Q8KY71
ID Q8KY71 PRELIMINARY; PRT; 86 AA.
AC Q8KY71;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium gastris.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M97-938;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253461; AAM69223.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9722 MW; BAF085B56BFD1DBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 36
Q8KY78
ID Q8KY78 PRELIMINARY; PRT; 86 AA.
AC Q8KY78;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Japan;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253454; AAM69216.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9695 MW; A5D2B8AA472C35F3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

Q8KY77
ID Q8KY77 PRELIMINARY; PRT; 86 AA.
AC Q8KY77;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium marinum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC927;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253449; AAM69211.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 38
Q8KY79
ID Q8KY79 PRELIMINARY; PRT; 86 AA.
AC Q8KY79;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=China;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253453; AAM69215.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29
```

```
RESULT 37
Q8KY80
ID Q8KY80 PRELIMINARY; PRT; 86 AA.
AC Q8KY80;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium marinum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC927;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253449; AAM69211.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 38
Q8KY79
ID Q8KY79 PRELIMINARY; PRT; 86 AA.
AC Q8KY79;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=China;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253453; AAM69215.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29
```

```
RESULT 39
Q8KY77
ID Q8KY77 PRELIMINARY; PRT; 86 AA.
AC Q8KY77;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Surinam;
RA Shamputa I.C., Bastian I., Ponteyne P.-A., Portael F.;
RT "detection of a gene encoding an alpha-crystallin homolog in numerous
RL mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253455; RAM69217.1; -.
DR InterPro: IPR002068; HSP20.
DR Pfam: PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON TER 1
FT NON TER 86
SQ SEQUENCE 86 AA; 9778 MW; A5D2B8B09DEC35F3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
Db 23 RYELRAE 29

RESULT 40
Q9BK93
ID Q9BK93 PRELIMINARY; PRT; 91 AA.
AC Q9BK93;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Polythreonine-rich glycoprotein (Fragment).
OS Cryptosporidium sp. 'muskrat 603'.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=138541;
RN [1]
RP SEQUENCE FROM N.A.
RC NEDLINE=21141828; PubMed=11229905;
RA Perz J.F., Le Blancq S.M.;
RT "Cryptosporidium parvum Infection Involving Novel Genotypes in
RL Wildlife from Lower New York State.";
RL Appl. Environ. Microbiol. 67:1154-1162(2001).
DR EMBL; AF303055; AAK20279.1; -.
FT NON TER 1
FT NON TER 91
SQ SEQUENCE 91 AA; 10009 MW; 794C7FEEDAC9D19C CRC64;

Query Match 0.9%; Score 7; DB 5; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLKPIV 10
Db 24 TTLKPIV 30
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:57:37 ; Search time 22 seconds
(without alignments)
1457.801 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLPKIVLSILLINPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310958 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1391	34.5	704	4	US-09-328-352-5144
2	685	17.0	718	4	US-09-252-991A-29432
3	306	7.6	741	4	US-09-328-352-5898
4	246	6.1	753	4	US-09-252-991A-17987
5	232.5	5.8	789	4	US-09-252-991A-27011
6	227	5.6	652	1	US-08-765-081-6
7	227	5.6	652	3	US-09-098-082-6
8	216	5.4	713	3	US-08-628-434-4
9	203	5.0	725	3	US-08-425-843-2
10	202	5.0	713	3	US-08-628-434-2
11	198	4.9	617	4	US-09-252-991A-30921
12	193	4.8	707	4	US-09-328-352-4834
13	189	4.7	715	3	US-08-425-843-7
14	188.5	4.7	663	1	US-08-765-081-7
15	188.5	4.7	663	3	US-09-098-082-7
16	182	4.5	718	5	PCT-US95-06994-7
17	180.5	4.5	914	2	US-08-487-890A-11
18	180.5	4.5	914	2	US-08-478-435-11
19	180.5	4.5	914	2	US-08-337-483-11
20	180.5	4.5	914	2	US-08-478-373-11
21	180.5	4.5	914	3	US-08-474-671-11
22	180.5	4.5	914	3	US-08-483-577A-11
23	180.5	4.5	914	3	US-08-897-438-11
24	180.5	4.5	914	4	US-08-637-654-11
25	180.5	4.5	914	4	US-08-649-518-11
26	179	4.4	756	4	US-09-252-991A-30809
27	174	4.3	761	4	US-09-328-352-5942

28	173	4.3	723	4	US-09-252-991A-18279	Sequence 18279, A
29	170.5	4.2	706	4	US-09-252-991A-32448	Sequence 32448, A
30	169	4.2	595	4	US-09-252-991A-25815	Sequence 25815, A
31	165.5	4.1	1245	4	US-09-252-991A-30935	Sequence 30935, A
32	164.5	4.1	643	4	US-09-328-352-5146	Sequence 5146, Ap
33	163	4.0	913	1	US-08-487-890A-5	Sequence 5, Appl
34	163	4.0	913	2	US-08-478-435-5	Sequence 5, Appl
35	163	4.0	913	2	US-08-337-483-5	Sequence 5, Appl
36	163	4.0	913	2	US-08-474-671-5	Sequence 5, Appl
37	163	4.0	913	3	US-08-483-577A-5	Sequence 5, Appl
38	163	4.0	913	3	US-08-897-438-5	Sequence 5, Appl
39	163	4.0	913	3	US-08-637-654-5	Sequence 5, Appl
40	163	4.0	913	4	US-08-649-518-5	Sequence 5, Appl
41	163	4.0	912	1	US-08-487-890A-7	Sequence 7, Appl
42	161.5	4.0	912	1	US-08-487-890A-9	Sequence 9, Appl
43	161.5	4.0	912	2	US-08-478-435-7	Sequence 7, Appl
44	161.5	4.0	912	2	US-08-478-435-9	Sequence 9, Appl
45	161.5	4.0	912	2	US-08-478-435-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

US-09-328-352-5144
; Sequence 5144, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5144
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5144

Query Match		34.5%	Score 1391;	DB 4;	Length 704;
Best Local Similarity		40.6%	Pred. No. 5.2e-125;		
Matches		315;	Conservative 121;	Mismatches 222;	Indels 118; Gaps 24;
Qy	9	IVLSIL-LINTPLLAQAHETEQVGLTETVTVGKSRPRATSGLLHTS---	TADKTIISGD	64	
Db	21	LTLSILAAISIPVFAAENVEK--LETTIRI--KAHP-----	LEQTSKDFAVADTVDDQK	71	
Qy	65	TLRQAVNLGDALDGVPGIHASOYGGASAPVIRGOTGRRIKVLNHHGETGMADFS	PDH	124	
Db	72	HLTEGATIGDALNSEVGIYANQFAGSSEPVIRGQDGPVKVLQNSSENVDTLS	SPDH	131	
Qy	125	AIWVDTALSOQVIRLGRPVTLTYSSNGVAGLVADGKIPKMPENGVSGLRLSSG	N	184	
Db	132	AVTVDPVLAQVEVIRGPTLLFGAGTVGLVNVINDNKIPTQMPENGVEQVGLRV	NTGS	191	
Qy	185	LEKLTSGINIGLGNKPVFLHTEGLYKSGDYAPRY-----	RNLKRLPDSPPRANGQ	240	
Db	192	DEKLASAGTVGLGSQVALREGLTDANNYIAPNTIHEGEKERRVDNT---	FAQDSDVN	248	
Qy	241	V-LGWRKFRYR-----TVSDRRDQGLPAHSHEYDDCHADIWQKSLINKRYLQ	LYPHL	294	
Db	249	VGLSW---IYDRGYTGISYSNREDQGLFCHSHEYSCSA-----	HL	287	
Qy	295	LTRDVDYDNPGLSCGFHDDD-----DAHAHA-----	HNGKPTDLRNKRYELRAE	WQPF	345
Db	288	-----GGRPHLHCAHEHDEHEEGEHAHAHEHEGPGWDLKSEYDFKTELND	PP	340	
Qy	346	PGFEALRVHLNRNDYHHDKAGDAVENFNQONARIELRHOPIGRLKSGMWQVY	LGOK	405	
Db	341	AGFQKLRAQASYTDYQHDEIEEGTATIRFQNGYDRIELVHNPFIADWEGVIGTQ	-LGQQ	399	

QY 406 SSALATSEAVKQPMMLDNKVOHYGFFGVQEQANWNTFLEGGVRVEKQKASIRVDKALID 465
Db 400 K--LNLGTGE--EAFMAPTTTTKWVSFALEHKQWQDVHFEISAPADQGE-----ID 445
QY 466 RENYKQKPLDCAHQRTARSFALSGNYFTTPOHKLSTASHQERLPSTQBYLAHGKHVA 525
Db 446 VDNKSKQDFG-----SAFSYAGANMEFAFNKLSFVASHQERLPLAQLYANGAHTA 499
QY 526 TMTFVGNKHLNKRNSNIELALGYEGDRWQYNLALYRNFQNYIYAQTLDNGRGPXIE 585
Db 500 TMTYELGNDQLSKNSNNVELGLHFDNDKLDYHLHVYHNFDDYIYAQTLDNRYK 553
QY 586 DQSEMLKLYNOSGADFYGAEBEIVFKPPTRYRIGVSGDYVRGRKLNPLSLPGRDAYGN 645
Db 554 ---DRLVQYTQKARFYGAEBEIGYQITPMYKISAFGDYVRGKI-----DAEG- 599
QY 646 RPIAQQDONAPRVPAARLGFHLKASLTDRIANLDYRVPFAONKLAARYETRTPGHMLN 705
Db 600 ---NAPRIAPAGRLGTVDADFGDGFSGSAEYHVFNQDKIAAYETETEGYNMLN 650
QY 706 LGANYRNTRYG---EMWVYKADNLLNOSVYAHSSFLSDTPQMGRSFTGGVNVKX 758
Db 651 LGWAY--SCQYGAKTDIRYMKANLLDDTVIQAHSFLSNIFQVGRNFTVGVDPSF 704

RESULT 2
US-09-252-991A-29432
; Sequence 29432, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29432
; LENGTH: 718
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29432

Query Match 17.0%; Score 685; DB 4; Length 718;
Best Local Similarity 27.3%; Pred. No. 6.6e-57;
Matches 213; Conservative 120; Mismatches 319; Indels 122; Gaps 25;
QY 6 LKPIVLSILLINTPLL-AQAHETEQSVGLTETTVVGVKSRPRATSGLLHTSTASDKIISGD 64
Db 46 LTPALACLIVSGETLGDGRPSLPSQVITANPLGNESFATPS-----SVLEGD 95
QY 65 --TLRQKAVNLGALDGPVGIHASQYGGGASAPVIRGQTRIKVLNHHGETGDMADFP 122
Db 96 ELTLRQXG-SLGTGLNGLPVSSTYFGPGASRPVIRGMDGRIILLNGLGALDASSLSY 154
QY 123 DHAMVDTALSQQVEILRGVPTLLYSSGNVAGLVVDVADGKIPKMPENGVSGLRLSS 182
Db 155 DHAVPEDPNSVERLEVVRGPAALLYGNAITGGVNSFDNRIPSE-PVDGTHSGELRYGG 213
QY 183 GNLEKLTSGGINTGLGNFVLHTEGLYRKSGDYAVPRYNLKLPSDPRPF--ANGQHEA 240
Db 214 ANTRRSKGALEAGDG-NFALHVDAAASRFNDVIRIPGIAHSS---QRQIDGDTQKHV 268
QY 241 VLGRKRYRRTYSDRRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQYPLHTEEDV 300
Db 269 -----QNSDGRQDGAAGVGSVHWEHYAGLSY-----SGYDS 300
QY 301 DYDNPGLSCGFHDDDDAHAAHNGKWPIDLNRKRYELBAEWKQFPFGFPAALRVHLNRNDY 360

Db 301 NYGSPA-----EDDVR-----LKMQQODRYAFASEIRDLGEPFTSLKLDAAATKY 344
QY 361 HHDEKAGDAVENFNQOTQWARIELHOPILGRKSGWQYLCQKSSALSATSEAVKQPM 420
Db 345 EHKEIEDGETGTTFKNEGYEGRIEARHPLGPNVGVGAQFANSRFSALG-----EAF 398
QY 421 LLDNKVQHYSGFFGVEQANW---DNFTLEGGVRVEKQKASIRVDKALIDRENYKQPLD 477
Db 399 VPHTETDSALFALEE--WKLSDRLDLSFGARLE---HTRVDPDAKNERPAEND----- 448
QY 478 GAHQRTARSFALSGNYFTTPOHKLSTASHQERLPSTQBYLAHGKHVAINTFVGNKHLN 537
Db 449 GSQSFRTTSGLSSTGAVYKLTPIWLSAATLSYTERAPTFYELYANGPAAATGTVEVGADAD 508
QY 538 KERSNNIELALGYEGDRWQYNLALYRNFQNYIYAQTLDNGRGPK-----STEDDSEMK 591
Db 509 KEKAVSTDLALRFDNGVHGSVGVFFSRFSNYI--GLLASGRHRNEGEVVAASDDDEALP 566
QY 592 LVRYNOSGADFYGAEBE---IYFKPTPRYRIGVSGDYVRGRKLNPLSLPGRDAYGNRP 647
Db 567 EYLYSGVRADFYGVEAQDRIHLLSPYGNPDLESLGDTYTRAKND----- 611
QY 648 FIAQDDQONAPRVPAARLGFHLKASLTDRIANLDYRVPFAONKLAARYETRTPGHMLN 707
Db 612 ---TGEPLPRIAPRLNTALIWEL--QQQARVDVDEHAASQHRVPEELSTDGY--TTLG 664
QY 708 ANYRNTRYGWNW--YVKADNLLNOSVYAHSSFLSD--TPQMGRSFTGGVNVKX 758
Db 665 ASLGYNFDLGEBSRLAPVKGTNTLTNTQTVRYASSILRDRVPAAGRGIEBAGVKVAF 718

RESULT 3
US-09-328-352-5898
; Sequence 5898, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5898
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5898

Query Match 7.6%; Score 306; DB 4; Length 741;
Best Local Similarity 20.5%; Pred. No. 2.5e-20;
Matches 163; Conservative 125; Mismatches 337; Indels 172; Gaps 32;
QY 59 KIIISGDTLRQKAVNLGALDGPVGIHASQYGGGASAPVIRGQTRIKVLNHHGETGDM 118
Db 16 KVBSNT-----IGDALKNITGISTGFPNAGAPVIRSLGCRNVRGVENGEBFNGMN 68
QY 119 DFPDPAHMDVLTALSQQVEILRGVPTLLYSSGNVAGLVVDVADGKIPKMPEN----- 170
Db 69 AFSGHINIPFPDPIFEKVIYVKNKNTDNIYGNATGGSVQIESGLIPKXIEKPNKLDIVF 128
QY 171 -----GVSG-----ELGLRLSS-----GNLE-KL-----TSGGINI 195
Db 129 RKGFNDFVKGFNFNINDKNWSTNIYSEVEISSYKIPGNSKAKLCEQIFNSSGGINS 188
QY 196 GLGKNF-----VLHTEGLYRKSGDYAVPRYNLKLPSDPRPFANGHVRVGRKPYR 250
Db 189 ALAASCQKDSRVQH---IYNKSQPYDKF--MTENPD---WADGDF-----SFYT 231
QY 251 -----RYSRRRQYGLP-AHSHEYDDCHADII--WOKSLINKRYLQYPLHTE 298
Db 232 DKPTSIWGGKTYINPKNPEYIPNTQNTIKKINTDVTNPFKLGNS-YAQ-----NE 283

```

299 QY DWVDYNP-----GLSC-----GFHDDDAHAHANGKP-WIDLKXKYEIURA 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
284 Db NIGFTTTFYFKFGFISADKKSEYGVGFSLQNSFADSETLPEVGKIDQNFVLNS 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
340 QY EWKQFPFGFEALRVHLNRNDVHDEKAGDAVENFNNQTNQRIELRHQPIGRLKSGWGV 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 Db NFIQDISAEBISLNFQLSNKSGEYVGTAKANEYKIDNQLIELIMKQSSFGLDGILGF 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 QY LQKSGSALSATSESAVKQPMLLDNKVQHVSYFGVEQANWDNTLGGVVRVEKQKASTRY 459
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
404 S-LKRNRETSQGTORYLPNVSTISKAJ-----FLOEBLNIKQFTLMTGYRFERIBHEL-- 455
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
460 QY DRALIDRENYKQPLDLAGHRQTAR-----SFALSGNWFYTPQKLSLTASHOER 510
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 Db -----QDNDFK-----LARASNRIENRYNLNQYFIGEYKVNFINLKVDYISER 504
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 QY LSTOELVYAHGHVATNTFVGNKHLNKRSENNITELALGYEGDORQVYNLALRYNRPGVY 570
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 Db APEVELYSSNNVHYVMAQEEGDQNLKPEKSKSLGMLFNWNNWYMLVGYQMDFNTM 564
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
571 QY YACTLNDGRGPKSIEDDEMKLVRYKNSGADFYCAEGBE1YF----XPTPRYRIGVSGDV 626
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
565 Db -----SHSGVAVQNRLLPLKYWKQTTDVTGKFELDLYDFNLAHIGNIKLGLADVF 617
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
627 QY RQRLKNVLPSPGOREDAYGNRPFIADDDQNAPRVPAARKLGFHLK---ASLTDRIDANLDVY 683
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
618 Db KKNKAVNPTDIR-----LANDGYVLPNMPTRNYGMFEWRNDSWKGKISS--IYV 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
684 QY ---RVFAONKCLARYETRTPGHMLNLGANYERNTRYGEMWNYVKADNLLNQSIVYAHSSPL 740
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
665 Db DEPRYLGRKVI--QEVPLSGYNLLEFKDIXKLKKNASFDLIFNGTLLNMBEARPONSPL 722
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
741 QY S-DTPQMGRSFTGGVNV 756
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
723 Db KYIAPLPGRAFOLGITM 739
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
US-09-252-991A-17987
; Sequence 17987, Application us/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

```

Query Match          6.1%; Score 246; DB 4; Length 753;
Best Local Similarity 21.2%; Pred. No. 1.6e-14;
Matches 184; Conservative 102; Mismatches 277; Indels 304; Gaps 42;

Qy      4  TTLXPVLSILLINTPLIAQAHTEQSVGLETTVVGKSRPRATSGLLHTSTASDKII-S 62
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      37  SSLSRLAIVAILGVSMQAQAEERKEGLG--TVTVGDMGLCEADQAVVQNHFGARSVVR 94
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      63  GDTLRQRAVNLGDALDGVPGIHASQYGGGASAPV-----IRGQTRRIKVLNHHGETGDM 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      95  REMLESGAQNVRDVLKYPGVQVDNNGTGGSDISLNVGVRLTSR----- 140
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      118  ADFSPDHAINVD-----TALS-----QQVEITLRGPEVTLLYSSGNVAGLVD 157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

141	--LSPSTVMIIDGVAANAPYQCPQLSMWPLSIGNLESIDVVRGAGSVRYGPNQVGGVIN	198
158	VADGKIPEKMPENGVSIGELGLL-----SSGNLEKLT-----SSGINIGLGNFVHTEGLYR	210
199	FYTRAIPKEF-----SGSIGTTIEHAGHGKWKLNQAFLLGGTGADNGLG--VALLYSGV--	249
211	KSGDVAVPRYRNKLRLPDSPRFANGQHVRLGMEKFRFYRRTYSRRRQYGLPAHSHEYD	270
250	KGADY-----RDGNNDIIDVLLKTHQLWLTDSQ-----LAANFHYI--	287
271	DCHADIIWOKSLINKRYLOLYPHILITEEDVDYDNGLSCGFHDDDDAHAAHNGKFWIDL	330
288	DAYADM-----EGGLTQAYD--DDPFQSV-----RDWDMF	316
331	RNKR-----YELRAEWKQPPFGFEALRVHLNR-----NDYHDEKAGDA	369
317	RGRKXDFSLKYTRQVDDLTQFEVLTVYSDSPRGSSIAARNLRTITSYPRDYHVFAVEPRV	376
370	VENFNQNTQNARIELHOPIELKSGWGVYLGQKSSALSATSSEAVKQPMLLDN-----	424
377	SRIFAGPTT-----QEVG-----IGRYLKE-----ANERASQIALVDNVPTVR	417
425	-----KVQHYFFGVGEQANWDMNFTLEGVVRVEKQKASIRYDKALI-----	464
418	PGSDGHTYDRTGGTEASAFYIDDKIDVGNWITTPGIRFEKIDSDWR--DRPVLGLNGKPV	476
465	---DRENYIKQPLDGLAHRQTARSPALSGNMYFFPOHKLSLTASHQELPSTOELYAHG	521
477	QEKRSKQDNEELPAL-----SLMYHLSDEW-----KLPAHYSESGSLQ--YFQLG	521
522	KHVATNTFVGNKHLNKNKRSNNIELALGYEDGRWQYNLALYRNRFNYIYAOTLNDGRGP	581
522	QGGSGNDTAAG--LPEKAKTYELCTRYDNGSWGGEITLF-----	559
582	KSTEDDSEMKLYRNGSGADFYCAEGSIYFKPTTPRYRIGVSGDY-----VRG	628
560	-YIDFDELQYVS--NDVGWNLGA-----TKHQIETSGHYDPAALDPRIDGLSVYVG	609
629	RLK-----NLPSPFGREDAYGNRPFFIAQDDQDNAPRVPAAARLGFHLKASITDRIDANL	680
610	SLTYTRATYEGDIPSPFKGRDLPLYSR-----QVATAGVRYEV-----DRWTYNL	653
681	DYRVFAQNKLARYETRTPGHMLNLGANYRN-----TRYGE-----WNWVVKAD	726
654	D--AFAQSM-----QRAFLSTDSOG--NPTHNYITEPSADGGYGDIPGYVTWNARGYD	704
727	-----NLLNQSVYAHSS	738
705	FGPQASNLKLGKVGKLNLFDKQYFTRSS	731

RESULT 5
 US-09-252-991A-27011
 ; Sequence 27011, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: ARBUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

Db 323 STLV---APIGHEMLSPGVE--GKHESLEDKTSNKISSRTHISN-TQWAGFIEDEWALAE 376
Qy 441 NFTLEGGVVRVEKQKASIRYDKALIDRENYKQPLDLAGHROTARSPALSGNWTTPQHK 500
Db 377 QFRLTFGRLDHDK-----GSH-----FSPRVYGVNLDPLMT 411
Qy 501 LSLTASHQERLPSTOELAHGKHVATNTFEVGNKHLNKNERSNNIELALGYE-GDRWQYNL 559
Db 412 VKGGVSTGFRAQLREVTPDWQVSGGNIYGNPDLPQETSINKELSLMYSTGSLAASL 471
Qy 560 ALYENRF-----GNVIAQTLNDGRGPKSIEDSEMMLVRYNOSGADFYGAEGEI 609
Db 472 TAFHNDFKDKITRVACPNICTAGPNQWATP-----TYRVNIDEAETYGAEATL 521
Qy 610 YFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAVRPAARLGFHLK 669
Db 522 SLPTESVELSSSYTYTHSEQS-----GNFAGRELL-----QLPKHLEFNANLS 565
Qy 670 ASLTDRIID--ANLDY--YRVFAQNKLYETRTFEGHMLNLGANYRRNTRYGEMWYVKA 725
Db 566 WQTTDLRLNSWANLYRGKEMQPEGASNDDFIAPSYTFIDTGVYALTD-----TATIKA 620
Qy 726 D--NLLNOSV-YAHSSFLSDTPQMGRSFTGGVNVKF 758
Db 621 AVYNLFDOEVNAYEYGYVED----GRRYWLGLDIAF 652

RESULT 7

US-09-098-082-6
; Sequence 6, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli O157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:

; ORGANISM: Vibrio Cholerae
US-09-098-082-6

Query Match 5.6%; Score 227; DB 3; Length 652;
Best Local Similarity 21.6%; Pred No. 8.3e-13;
Matches 176; Conservative 108; Mismatches 310; Indels 222; Gaps 40;

Qy 1 MAQTLKPIVLSI---LLINTPLLAQ-AHETEQSVGLTFTVVVGVKSRPRATSGLLHTSTA 56
Db 1 MSRFNPSVSLSVTLGLMFSASAFADATKTDETM---VVTAAGYAQ-----VIQNAFA 51
Qy 57 SDKIISGTLQKAV-NLGDALDVGPGIHASQYGGASAPV-IRG-----Q 100
Db 52 SISVISREDLESRYRDVTDALKSPGVTVT---GGGDTTDSIRMGSGNYTLILVDGKRQ 109
Qy 101 TGRRIKVLNHGETGDMADFDPDHAIWVDTALSOQVEILRGVPTLLYSSGNVAGLVADVAD 160
Db 110 TSQRTRP-NSDQPGIEQGLPPLQAI-----ERIEVIRGPMSTLYGSDALGGVNIIT 161
Qy 161 GKIPKMPENGVSGBELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYKSGDYAVPRY 220
Db 162 RKDQQW-----SGNVQLSTVVQEN-----RASGD----- 186
Qy 221 ENLKELPDSRRFANGQHRVGLGWRKRPYRRTYDRDOYGLPAHSHEYDDCHADIIWOK 280
Db 187 -----EOSANFFVTGPLSDALS--LQYVGTQTORDEDEI-----EHGYGDKSL-----R 228
Qy 281 SLINKRYLQLYP--HLLTEEDVD---YDNPGLS-----CGFHDDDDAHAH---AH 322
Db 229 SLTSKINYQLNPDHOLQLEAGVSAQDRENNVKSQSSGCRGCTCNTDQYRNHVAVSH 288
Qy 323 NGKQWIDLNRKRYELRAEWKQFPFGFEALRYHLNRNDYH--HDEKAGDAVENFFNNQTN 380
Db 289 QG-----DWQ-----GVGQSDTYLQVEENTNKSREMSIDNTVPK 322
Qy 381 ARIELRHOPIGRLKSGWQVYLQKSSALSATSEAVKQPMLLDNKVQHYSPFGEQANWD 440
Db 323 STLV---APIGHEMLSPGVE--GKHESLEDKTSNKISSRTHISN-TQWAGFIEDEWALAE 376
Qy 441 NFTLEGGVVRVEKQKASIRYDKALIDRENYKQPLDLAGHROTARSPALSGNWTTPQHK 500
Db 377 QFRLTFGRLDHDK-----GSH-----FSPRVYGVNLDPLMT 411
Qy 501 LSLTASHQERLPSTOELAHGKHVATNTFEVGNKHLNKNERSNNIELALGYE-GDRWQYNL 559
Db 412 VKGGVSTGFRAQLREVTPDWQVSGGNIYGNPDLPQETSINKELSLMYSTGSLAASL 471
Qy 560 ALYENRF-----GNVIAQTLNDGRGPKSIEDSEMMLVRYNOSGADFYGAEGEI 609
Db 472 TAFHNDFKDKITRVACPNICTAGPNQWATP-----TYRVNIDEAETYGAEATL 521
Qy 610 YFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAVRPAARLGFHLK 669
Db 522 SLPTESVELSSSYTYTHSEQS-----GNFAGRELL-----QLPKHLEFNANLS 565
Qy 670 ASLTDRIID--ANLDY--YRVFAQNKLYETRTFEGHMLNLGANYRRNTRYGEMWYVKA 725
Db 566 WQTTDLRLNSWANLYRGKEMQPEGASNDDFIAPSYTFIDTGVYALTD-----TATIKA 620
Qy 726 D--NLLNOSV-YAHSSFLSDTPQMGRSFTGGVNVKF 758
Db 621 AVYNLFDOEVNAYEYGYVED----GRRYWLGLDIAF 652

RESULT 8

US-08-628-434-4
; Sequence 4, Application US/08628434
; Patent No. 6265567
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule
; NUMBER OF SEQUENCES: 4


```
Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;
QY 11 LSILLINTPLAQAHETEOSVGLTWTVVGKSRPRATSGLLHTSTASDKIISGDTLRQKA 70
Db 6 LSLAIAATT--LVTANALAQVELDSINVIATRDP---SREAYTPEKQSK--DSLISKQA 57
QY 71 VNLGDALDGVPGFIHASQYGGG---ASAPVIRGOTGR-IKVLNHHGETGMADSPDH-- 124
Db 58 TSVAALLEDIPNDVIR--GGSRSIAQPNIRGLSDNRVQVIDGVRFDLA-----HRG 110
QY 125 AIMVDTALSOQVEILRGVPTLLXSSGNVAGLVADVADGKIPEKMPENGVSGLGRLSSG- 183
Db 111 SYFLPMSLIQIEIVIKPSSSLGSGALGGVAMRTPNALDLLKN--DKFGVKIRQGY 167
QY 184 -NLEKLTSGGINT-GLCKNFVLHTEGLYKSGDYAVPRYRNLRKLPDPRPFANGQHRV 241
Db 168 QTANNLSERDVSVFAANDKFDVLISGFYNNADNLRTGKGNLKN--TAYKQFGG---LAK 222
QY 242 LGWRKRFYRET-YSDRRDQGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDV 300
Db 223 FGQINDANVELSHRETRFKQTAPSN-----NEVENELTNEQITDQIR 266
QY 301 DYNPGLSCGFHDDDDAHAAHAKPWIDLRNKRYELRAEWKQFPFGFEALRYHLNRNDY 360
Db 267 EFHKP-----NNGSP-----PRAKPSQEEF-----YSGVKTRFGSVSY 299
QY 361 HHDEKAGD--AVENFR---NNQTONARIELRHQPIGRLKSGWGVOYLQKSSALSATSEA 415
Db 300 LTDQIQDQSTVFNYITPDNLYNTHIALYNNKT-----IEKQKVSQVKDQTKL 351
QY 416 VKQPMLLDN--KVQHSYF-FGVEQANWDFTLBGGV-----RVEKQKA----- 455
Db 352 TTRGINLRNSELSSHISFVGVVDYMR-DKIRTERGTNNKDAQFRADPYNANSNTTGYILI 410
QY 456 -----STRYD-----KALIDRENYKQPLDLAGHROTARSAFALSGNWYF 495
Db 411 AHIPFEGEKLLSPSVRYDHYDTSSTVKYKDNHL-----SPATKLTWIV 455
QY 496 TPOHKLSLTASHOE--RLPSTQELYAHGKHVAT-----NPFVGNKHLNERSNNIE 545
Db 456 T--NWLDFTAQYNEAFAPSMQERFVSGSHFGTSILGRNEINKEF-VANPNLRPETAKNKE 512
QY 546 L-----ALGYBGDRQWYNALALYRNFYNYIAQTINDGRGPKSTEDDSEMKLVRYNOS 598
Db 513 ITANLHFDSLFKQGDKEFKIEATYFRNDVKDFINLKIFNDAKTNNASASA-----GA 564
QY 599 GADFYCAEGEIKYKPT-PRYR-----IGVSGDYVRGRL---KNLPSLPQREDAYGN 645
Db 565 GA---GANPNGALLPTKSQYQNTNARLSGIELQAOYQTERLTLFTNYGSGTKGDKDSG- 620
QY 646 RPFIAQDDQNAVRVPAARLGFHLKASITD--RIDANLDYRVFAQNKLAHYETRT-PGH 701
Db 621 -----EALSNIASKTGVGVNYALVKDKFTVGATVTTHYA--AQRVPVKDHSVTYPSY 670
QY 702 HMLNLGANYRRNTRYGEW 719
Db 671 ILTDLRAATYA--PLKGEW 686
RESULT 10
US-08-628-434-2
; Sequence 2, Application US/08628434
; Patent No. 6265567
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule and Vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
```

```
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,434
; FILING DATE: 05-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/418,964
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-3-P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-628-434-2
```

Query Match 5.0%; Score 202; DB 3; Length 713;
Best Local Similarity 21.8%; Pred. No. 2.5e-10;
Matches 187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;

```
QY 11 LSILLINTPLAQAHETEOSVGLTWTVVGKSRPRATSGLLHTSTASDKIISGDTLRQK 69
Db 8 LSLLSITLAAGFAHAENANVALDIVTVKGRQ-----GSKTRTNIV---TLQOK 55
QY 70 ----AVNLGDALDGVPGFIHASQYGGASAPVI-----RGQTGRRIKV-----LNHH 111
Db 56 DESTATDMRELLKEEPSI---DFGGNGTSGQFLTRGMGQNSVDIKVDNAYSQSIILYHQ 112
QY 112 GETGDMADSPDHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMP-- 169
Db 113 GR-----FIVDPALVKVSVQKG-----AGSAGIGATNGAIIAKTVDAQ 153
QY 170 ---NGVSGBELGLRLSSG--NLEKLTSGGINIGLKNFVLHTEGLY-----RKSGDYAVR- 219
Db 154 DLLKGLDKWGVRLNSGFAGNCGASGVFGKEGNF---DGLFSYNRNDEKDIYAGKG 209
QY 220 YRN---LKLPLDSP-----RRFANGQHRVAVLGMWRKFRYR--RTYSDRRDQYGLP 263
Db 210 FRNDNGGKTPYSALDKRSYLAKIGTTFGDDHRIVLSHMKDQHRGIRTV---REEFAVS 266
QY 264 AHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHN 323
Db 267 EKN-----SRITIKRQAPSIRE--TQSNNTLAYTKOLGFEKLDANAYV-- 310
QY 324 GRPWIDLRNRYELRAEWKQFPFGPEALRVHLNRNDYHHDEKAGDAVENFFNNQTNARI 383
Db 311 -----LEKKRYS--ADKX-----NGYAGNVKGNPNTRIATRSMMNFDS 348
QY 384 ELRHQPIGRLKSGWGVOYLQ--KSSAL-----SATSEAVKQPMLLDNKVQHSY 430
Db 349 RLAEQTL--LK--YGINYRHOEIKQPAFLNSFEIKDKEKATNEEKKKRENEKIAKAYR 404
QY 431 FGVGEQANW-----DNFTLEGVVEKQKASIRVDKALIDRENYKQPLDLAGH 480
Db 405 LTNPTKTDGTAYIEAIEHIDGFTLTGGLFRYDKVKTHTDKG----- 445
QY 481 ROTARSFALSGNW--YFTPOHKLSLTASHQ--ERLPSTOE-LYAHCKH-----VATNTPFV 531
Db 446 --TVSSSSLNPSFGVWQPREHWSFSASHNYAGRSPRLYDALQTHGKRGIIISADGT--- 500
```

Query Match	4.8%;	Score 193;	DB 4;	Length 707;	
Best Local Similarity	19.0%;	Pred. No. 1.8e-09;			
Matches 162;	Conservative 126;	Mismatches 312;	Indels 252;	Gaps 35;	
6	LKPIVLILLINTPLLAQAHEQTSVGLTETVVGKSRPRATSGLLHTSTASDIIISGDT	65			
9	LQPLWVAMLVSHSGELVFPAKSEKNDADNTILHSAPIVTTAQOQ----	64			
66	LRQ-----KAVNLGDALDGVPGIHASQYGGASAPVIRGQTORRIKVLNHHGET-GDMADF	120			
65	KQPIQVPVATGDADYLOSIQMFNSIQSGTGNGDVTFRGMFGSRKIILTDGTENIGACPNR	124			
121	SPDHAIWDTALSQOVELRGPVTLTYSSGNVAGLVDVADGKIPEKOMPENGVSSELGLRL	180			
125	MDAPTSYISPEYDRISVIKGPQTQYATNSAATV-----LFRQP-----	166			
181	SSGNLEKLTSGGINIGLGNFVLHTPEGLYRKSGDYAVPRYLNKLPLDPSRRFANGOHRA	240			
167	-----EKLT-----EKPYRQASVLLGSYGR-----DHNVEA	195			
241	VLGRKRFYRTYPSDRROQYGLPAHSHYDDCHADII--WOKSLINKRYLQIPHLLTE	297			
196	AVGDEKK-YIRLNAAR-----SESNSYQDGDGNTVPSAWKKM-----	231			
298	EDVDYDNPCLSGCFHDDDDAHAAHANGKPWIDLNRKYE-----	336			
232	-----NUDVALGTFPDENT-----WVEITGCKSDGESLYAGRENDGSGQFARESLG	276			
337	LRAEWKQPPGPFAL--RVHLNRNDYHHDEKA-----GDAVENFPNN	376			
277	LAPEKKNITDVIKKTEGOVNTSYNDHIMDNESLRTPPLVEMNHGGMTMLPNAVAMOVTR	336			

```

QY 377 OTQARIELRHOPIGRKSGV--QYLQKSSALSATSSEAVKQPMLLDNKVQHYSPFGV 434
Db 337 RTLNGRLAMTS--WKLSLITIGDQPNKGGMSPTWPSNNIPIYQDMRFQSYGAFGE 395
QY 435 EQANWDMFT--LEGGVR-----VEKOKA-----SIRYDKALIDRENYKQPLPLGAHR 481
Db 396 LGYQSNLNLKLVTVGLDRVTVVEDERADSQAKGFNTKLEKTLPSAFVWENQRDPL---- 451
QY 482 QTARSFALSGNMYFFPQHKLSTASHOERLPSTQELYA--HGKHVATNTFEVGNKHLNKE 539
Db 452 -----DFKSYIGLGVERNPDYWELEFSEHGNAGTNTF-----NGVNP 491
QY 540 RSNNIELALGYEDRQWYLN--ALYRNRFNYIYAQ-----TLNDGRGPKSIEDDS 588
Db 492 KT--LQDLGLFQQHQCALNIWASAYAGLVDDYILMNYHDHSHLHPNEHGCHSGHITPCA 549
QY 589 EMLVRYNSGADFYCAEGEIVFKPTPRYRIGVSGDYVRGLKNLPSLPCREDAYGNRPF 648
Db 550 K-----NVDATIAAGAEAGIGYQFTDRIQADLSAMYAWG--KN----- 584
QY 649 IAODQNAVRP--AARLGFHLKASLTDRIANLDYRVFAQNKLARVETRTPPGHHM--- 703
Db 585 -TTDDKPLQISPLEGLNIRYVA---DKYNLGLLWRAVAEQNRVSLHOGNIVGYDLKPS 640
QY 704 -----LNLGANYRNTRYGEWNTVVKADNL-----LNCVVVAHSSFLSDTP--QM 746
Db 641 KGFSTLSLNGSVNLRKXIDIVS-----VGIDNVFDKTYTTEHLNKAAGSAGFGFASERQFNNI 695
QY 747 GRSEFTGGVNVKF 758
Db 696 GRNYWVRMSMKF 707

```

RESULT 13

US-08-425-843-7

; Sequence 7, Application US/08425843

; Patent No. 6020154

; GENERAL INFORMATION:

; APPLICANT: Hansen, Eric J.

; APPLICANT: Cope, Leslie D.

; APPLICANT: Jarosik, Gregory P.

; APPLICANT: Hanson, Mark S.

; TITLE OF INVENTION: H. Influenzae HxuD and HxuC Genes, Proteins

; TITLE OF INVENTION: and Methods of Use

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/425,843

; FILING DATE: Concurrently Herewith

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Parker, David L.

; REGISTRATION NUMBER: 32,165

; REFERENCE/DOCKET NUMBER: AMCY:012/PAR

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (713) 789-2679

; TELEX: 79-0924

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 715 amino acids

; TYPE: amino acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-843-7

```

Query Match 4.7%; Score 189; DB 3; Length 715;

Best Local Similarity 19.9%; Pred. No. 4.5e-09;

Matches 163; Conservative 124; Mismatches 272; Indels 262; Gaps 45;

```

QY 11 LSIILLITPLAAAHETEQSUGLETVTVVSGSRPRATSGLLHTSTASDKIISGDTLRQKA 70
Db 6 LSLAITTT--LVATANALAQSVELDSINVIAITRDP---SRFAYTEKOSK---DSSLKQA 57
QY 71 VNLGDALDGVPGIHASQVGGG---ASAPVIRGQTGR--IKVLNHHGTGDMADSPDH-- 124
Db 58 TSVAALAEIDPNVDV--GGSRSLAQKPNIRGLSDNRVQVVIDGVQNFDLA-----HRG 110
QY 125 AIMVDTALSOQVEILRGPTVLLYSSGNVAGLVVDVADGKIPEKMPENGVSGLGURLSSG- 183
Db 111 SYFLPMSLIQIEVIGKPPSSLMGSGALGGVAVMRTPNALDLDLKN---DKFGVKIRQGY 167
QY 184 ---NLEK-----LTSGGINIGLGNFVLHTEGLYRKSQGVAVPRYR 221
Db 168 QTANLSEKDSVFAANDKFDVLISAFYNNADNRTGKGNKANTTA--YKOPGGLA----- 221
QY 222 NLKRLPDSRFA--NGOHRVAVLGMKRFYRTYSDDRDOYGLPAHSHEYDDCHADIW 278
Db 222 ---KFGWQINDANRVELSHRETRFKQT-----APSN----- 250
QY 279 QKSLINKRYLOLYPHLLTEEDVDYDNPCLSCGHDDDDAHAAHNGKPIDLRNRYELR 338
Db 251 ---EVENELTNEKIIDQINE---FHG-----SNNGLP-----QR 278
QY 339 AEWKPPPGF--EALRVHLNRNDYHDEKAGD--AVENFF---NNQOTVARIELRHQPIGR 392
Db 279 AKPSSETSAPYKVKTRFGSVSYLTDOQIPDOSTVENVYLTDPDYNLTHALYNNKT-- 336
QY 393 LKSGNGVOYLQKSSALSATSEAVKQPMLLDN--KVQHYSP--FGVEQANWDFLEGV- 448
Db 337 ---IEKQKQVSGVKDQTKLATRGINLNSSELSHISFVYGVGYMR--DKIRTERGTN 389
QY 449 ---RVEKOKA-----SIRYD-----KALIDRENYKQ 472
Db 390 NNDKFAEPYNANSNTTGVYLIHPIPLFGEKLLSPSVRYDHYDTSKTKYKDNHL- 447
QY 473 PLDGLGAHROTARSFALSNGWYFTPOHKLSTASHOE--RLPSTQELYAHGKHVAT- 526
Db 448 ---SPATKLTWKVT--NWLDLSAKYNEAFRAPSMQERFVSGSHFGATILG 492
QY 527 ---NTPEVGNKHLNKRSENNIEL-----ALGVEGRWQYNLALYRNFNRYIYAQTL 575
Db 493 ADHINKF--VANPNLRPETAKNKEITANLHFDLSLQKDDKFKIEATYFRNDVKDFINLKIF 551
QY 576 NDGRGPKSIEDDSMKLVRYNSGADFYGABGEIVFKPT--PRYR-----IGVSGDY 625
Db 552 NDANTSTS-----ANGA-----FLPTKSOYQWITNARLSGLELOAQY 588
QY 626 VRGL---KNLPSLPGREDAYGNRPFFIAQDDQNAAPRVPAARLGFHLKASLTD---RIDAN 679
Db 589 QTERLTLFTNYGSGTKGRDKSG-----EALSNTAAASKIGVADYALVKDKFTVGAT 639
QY 680 LDYRVFAQNKLARVETRT--FGHMLNIGANYRNTRYGEW 719
Db 640 ITHYA--AQHRVPKDHAVTYPYSILTVLRATYA--PLKGEW 676

```

RESULT 14

US-08-765-081-7

; Sequence 7, Application US/08765081

; Patent No. 5798260

; GENERAL INFORMATION:

; APPLICANT: Tarr, P.I., Balge, S.S., Besser, T.E., Vary Jr., J.C.

; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

```

; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli CirA protein amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia Coli
; US-08-765-081-7

Query Match 4.7%; Score 188.5; DB 1; Length 663;
Best Local Similarity 20.5%; Pred. No. 4.4e-09;
Matches 176; Conservative 110; Mismatches 270; Indels 301; Gaps 44;

Qy 6 LKPIV-----LSILLINTPLLAQHETEVSQVLETTVVGKSPRATSGLLHTSTASDKI 60
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 61 ISGDTLCKAV-NLGDALDGVPGTHASQYGGGASAPVIRG-----QTGRRIRK----- 106
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 54 ITQEDLQKPKQNLKQVLKEVPGVQLTNEGDNKGVSIKGLDSSYTLILVDGKVSNRNA 113
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 107 VLNHHGTGDMADSPDHAIKQVLEILGQVETLLYSNGVAGLVVDVADGKIPEK 166
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 114 VFRHN-----DFDLNW-IPVDSI--ERIEVVRGPMSSLYGSDALGVVNIITKIGOK 163
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 167 MPENGVSGLGL-----RLSSGNLEKLTSGINIGKFNVLHTEGLYKRSQGVAV 217
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 164 W-----SGTVTVDTTIEHRDRGTNGQFTSGPLDGV-----LGM 201
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 218 PYYRNL-KRLPDSPPRFA---NGQHRVAVLWGRKGFYRTYSRRDQYV-LPAHSHEY--- 269
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 202 KAYGLAKREKDDQNTTDTGTGTPTIEGFSR-----DGNVEFAMTPNQNHDFTAG 254
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 270 -----DCHADIWQSLINKRY-----LQLYPHLLITEEDVDYDNPGLSCG 310
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 255 YGFDQRQDSLSLKNLERQYVSNHGRWDYGTSLKYY-----GEKVENKPNQNSPP 309
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 311 PHDDDDAHAAHAGKPEIDLR--NKRVELRAEWKOPFFGFEALRVHLNRNDYHDEKAGD 368
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 310 I-----TSESNTVDGKYTLPLTALNQFLTVGGEW-----HD-KLSD 345
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Qy 369 AVENFFNNQTONARIELRHQPIGRILKGSWGQVYLGKSSALSATSEAVKQPMLLDNKQVH 428
Db 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

; 346 AV-----NLIT-----GTSSKTSAS-----QY 362
; 429 YSPFGEQANWDMFTLEGGVVRVKQKASIRYDKALIDRENIYKQPLDPLCAHROTARSPA 488
; 363 ALFVEDEWRIFEPLALTGTGRMD-----DHETYGEHWSPRAYLVYNATDITVT 409
; 489 LSGNMYETPOHKLSTASHOERLPSTOELVAGKHVATNTE-----VGNKHLNKRSSNN 543
; 410 VKGW-----ATAP-----KAPSLLOL-----SPDWTNSCRACKIVGSPDLKPTSES 454
; 544 IELALGYEGDR-W-----QYNLALYRN-----RFGNYIYAOTLNDGRG 580
; 455 WELGLYTWBEGWLEGVESVTVFRNDVKDRISRTSDVNAAPGYQNFVGFETGANGRR 514
; 581 PKSTEDDSEMLVRYNQSADFYGAGEIYFKTPPRYRIGVSGDYVGRKLNPLSLPGRE 640
; 515 IPVFS-----YINVNKNQCVETELKIPNDEWKLSTINTYNDGR----- 555
; 641 DAYGNRPFIADQDONAPRVPAARLGFHLKASLTDRIDANLDYY---RVFAQNKLARYETR 697
; 556 -----DVSNGENKPLSLPFLH---ALED-----WSFYVSGHYTGOKRADSAATAK 597
; 698 TPGHML-NLCA-----NYRNTFYGEWNYVYKADNLLNQSVYAHSSFLS 741
; 598 TPGGYTINWTGAAMQVTKDKLRAGVLNLGDKTANGTLDW--KPD--LSRDDYSYN--- 649
; 742 DTPQWGRSFTGGVNVKF 758
; 650 ---EDGRYFNAVDIRF 663

RESULT 15
US-09-098-082-7
; Sequence 7, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Parry, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid

```


QY 577 ---DGRGPKSIEDSEMKLVRYN-QSGADFYGABGEIYF----- 611
 Db 715 KNGTGKANYGHNAQNAKLGVNITAQDLDFNLGKWRIPYGYATFAYNRVKKVQDKINAG 774
 QY 612 ---KPTPRYRIGVSGD-----YVRGRLKNLPSLPQREDAYGNRPF 648
 Db 775 LASVSSYLFDAIQPS-RYIIGLYDHPSTNGINTMTQSKAKSONELLGKR-ALGNN-- 830
 QY 649 IAQDDQNAFRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLYARTETPGHHMLNLA 708
 Db 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYTMANK-----NIMLRGI 865
 QY 709 NYRNRTRYGEMWYVK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGGVNVKE 758
 Db 866 YNLFNRYVTWEAVRQTAQAVNQHVNGSVTRYAAS-----GRNYTLTLEMKF 914

RESULT 18
 ; US-08-478-435-11
 ; Sequence 11, Application US/08478435
 ; Patent No. 5922323
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/478,435
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/337,483
 ; FILING DATE: 08-NOV-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 914 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-478-435-11

Query Match 4.5%; Score 180.5; DB 2; Length 914;
 Best Local Similarity 19.7%; Pred No. 4.5e-08;
 Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

QY 5 TLKPI-----VLSILLINTPLLAQAH---ETEQSVG-----LETWVVG-KSRP 44
 Db 2 TKKPYPRLSIIISCLLISCVVKAETQIKOTKEAISSEVDVTQSTEDSELETISVTAEKIR 61
 QY 45 RATSGLLHTSTADSKII--SGDTLRQKAVNLGDALDVGPIHISQYGGGASAPV-IRGT 101
 Db 62 RQNEV-----TGLGKIILKTSISISREQVNLTRDLTRYDFGISVVEQGRGASSGYSIRGMD 117
 QY 102 GRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILFPGVTL--YSSGNVAGLDVA 159
 Db 118 RNRV-----ALLVDGLPQTQSVVQSPINVARSGYSGTGAINIEIYE 158
 QY 160 DGKIPKMPENGVSGLGLRLSSGNLEKLTSGGINIGL-KNFVLHTEGLYR----- 210
 Db 159 NVKAVE-ISKGGSSSEYNGALAGSVTFQSKAADILEGDKSWGIQTKNAYSSKNKGFT 217
 QY 211 -----KSGDY--AVPRYRNLKRLPDSPPRFANGQHR-AVLGWRKRFYRRTYSRRDOY 260
 Db 218 SLAVAGKQGGPDGVALYTORN-----SIETQVHKDALKGVS--YHRLIAPEDQ- 265
 QY 261 GLPAHSHEYDDCHADLIWQKSL-INKRYLQLYPHLLTEEDVY-----DNP----- 305
 Db 266 --SAYFVMDCEPKPDDYNSCLPFAPKRPAILSSQRETSVSDYTGANRIKPNPMKYESQS 323
 QY 306 -GLSCGFHDDDDAHAAHNGKWIIDLNRKYELRAEMKQPPGFEALRVHLNRNDYHDE 364
 Db 324 WFLRGGIHFSEQ---HYIGG--IFFTQOKFDIR--DMTFPAY-----LRSTE 364
 QY 365 KAGDAVENFFNQTONA--RIELRHQPIGR-----LKGSGVQVYLGQKSSA 408
 Db 365 KRDDSSGSFVPKQDYGAYQRIE--DGRGVNYASGLYFDBHHRKQRVGIEVIYENKNK 419
 QY 409 LSATSEAV---KOPMLDNKVOH-----YSPFGVE----- 435
 Db 420 AGIIDKAVLSANQOIIILDSYNOHTCHSLYPNPSKNCRPTRDKPYSYSHSDNRVYKEKN 479
 QY 436 -----QANW-----DNFT-----LEGVVRVEKQKASIRYDKALIDR 466
 Db 480 MLQLNLEKKIQNWLTQIVNLFGLDFDFTSALQKRDYLTFRVTATAKSI--EKANETR 536
 QY 467 ENYKQ-----PLPDIG-----AHRQTARSAFALSGNWFYFPQHKLSL----- 503
 Db 537 RNYGKKQPYLYPKPTVGVVQDHCYKNSNRYRDCVKRLIKGKNYFFAARNMALGKTV 596
 QY 504 -----TASHOE-----RLPSTQEL 517
 Db 597 DLGLGIRYDVSRTKANESIIISVCKPKNPSWNTGIVIKPTWLDLSYRLSTGFNPSFAEM 656
 QY 518 YAHGKHVATNTFVGNKHLNKNRSNIIELALGYEGDRWQYNLALYNNRFGNYI-YAQTJN 576
 Db 657 Y--GWRYGNNSEVYVGVKPKPETSRLNQBFGLKALGDFGNIEISHFSNAYRNLIAPAEELN 714
 QY 577 ---DGRGPKSIEDSEMKLVRYN-QSGADFYGABGEIYF----- 611
 Db 715 KNGTGKANYGHNAQNAKLGVNITAQDLDFNLGKWRIPYGYATFAYNRVKKVQDKINAG 774
 QY 612 -----KPTPRYRIGVSGD-----YVRGRLKNLPSLPGRBDAYGNRPF 648
 Db 775 LASVSSYLFDAIQPS-RYIIGLYDHPSTNGINTMTQSKAKSONELLGKR-ALGNN-- 830
 QY 649 IAQDDQNAFRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLYARTETPGHHMLNLA 708
 Db 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYTMANK-----NIMLRGI 865
 QY 709 NYRNRTRYGEMWYVK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGGVNVKE 758
 Db 866 YNLFNRYVTWEAVRQTAQAVNQHVNGSVTRYAAS-----GRNYTLTLEMKF 914

```

RESULT 19
US-08-337-483-11
; Sequence 11, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-337-483-11

Query Match 4.5%; Score 180.5; DB 2; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI---VLSILLINTPLLAQAH---ETEQSVG-----LETVTVG-KSRP 44
Db 2 TKRPFRLISLCILSYKVAETOSIKDTKEALISSEVDTSQSTSELETISVTAEKIRD 61
Qy 45 RATSGLLHTTASDKII--SGDTLRQKAVNLGDALDGVPGIHASQYGGGASAPV-IRGQT 101
Db 62 RKDNEV---TGLKIIKTSESISREQVLNIRDLTRYDPCISVVEQGRGASSYSIRGMD 117
Qy 102 GRRIKVLNHCETGDMADFDPDHAIWDTALTSQVEILRGPTVLL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPQTSYVQSPVLRVARGSGTGAINETYE 158
Qy 160 DGKIPEKMPENGVSGLGLSLSSGNLEKLTSGGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKGGSSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSSKNKGFTH 217
Qy 211 -----KSGDY---AVPEYRNLEKLPDPRFANGQHR-AVLGWRKPFYERTYSDRDQY 260
Db 218 SLAVAGKGGGEGVAYTYQRN-----SIETQVHKDALKGVQS--YHRLAKPEDQ- 265
Qy 261 GLPAHSHRYDDCHADIIWMKSL-INKRYLOLYPHLLTEEDVDY-----DNP----- 305
Db 266 --SAFVNVQDCEPKPDDYNSCLPFAKRAPIILSSQRETYSVDYTGANRIKPNPMKYESQS 323

```

```

Qy 306 -GLSCGFHDDDDAHAAHNAHNPWIDLNKRYELAEWKPFGPEALRVHLNRNDYHDE 364
Db 324 WFLRGYHFSEQ---HYIGG--IFEFTQOKFDIR---DMTPPAY-----LRSTE 364
Qy 365 KAGDAVENFNNQTONA--RIELRHQPIGR-----LKGSGWGYOYLQKSSA 408
Db 365 KRDDSSGSFYPKQDYGAYQRIE---DGRGVNYSGLYFDEHHRKQVRVGLYEIVENK 419
Qy 409 LSATSEAV---KQMLLNDKVKH-----YSFGVE----- 435
Db 420 AGIIDKAVLSANQOIIILDSYMOHTHCSLYPNPKNCRPTRDKPYSYSHSRNVYKEHN 479
Qy 436 -----QANW-----DNFT-----LEGVVRVEKQKASRYDKALIDR 466
Db 480 MLQNLLEKIIQONWLTHQIVFNLGFDFTSALQHKDYLTRRVATAKIS---EKANETR 536
Qy 467 ENYKQ-----FLPDIG-----AHRQTARSPALSGNWYFTPOHKLSL----- 503
Db 537 RNYGKKQPYLYPKPTVGVVQDHCDYKGNSSNYRDKVRLINGKNYFFAARNMALGKYV 596
Qy 504 -----TASHOE-----RFPSTQBL 517
Db 597 DLGLGTRYDVSRITKANESTISVGEKFNFSWNTGIVIKTEWLDLSYRLSTGPRNFSPEM 656
Qy 518 YAHGKHVATNTEFVGNKHLNKNERSNNIELALGYEDRWQYINLALYRNFGNVI-YAQTIN 576
Db 657 Y--GWRYGNNSEVYVVKPKPETSRLNQEFGLAKGDFGNIETSHFSNAYRNLIAPAEEL 714
Qy 577 ---DGRGPKSIEDSEMKLVRYN-OSGADFYGAEGEIVF----- 611
Db 715 KNGTGKANYGYHNAQNAKLGVGNITPAQLDFNGLNKKELPYGWYATPAYNRVVKVQDKINAG 774
Qy 612 -----KPTPRYRIGVSGD-----YVGRKLNPLSPDGRDAYGNRPF 648
Db 775 LASVSSYLFDALQPS-RYIIGLYDHPSTNWTGINTMTQSKAKSONELLGKR-ALGNN-- 830
Qy 649 IAQDDQNAVPAPARLGFHLKASLTDRIDANDLYRVFAQNKILARVETTPGHHMLNIGA 708
Db 831 -SRDVKSTRKLTRA---WHI-----LDVSGYTWANK-----NIMRLGI 865
Qy 709 NYRNRTRYGEWNVYK-ADNLLN--QSVYAHSSFLSDTFQMGSRSTGGVNVKPF 758
Db 866 YNLFNYEYVTVWEAVRQTQAQAVNQHVQSVTRYAAS----GRNYTLTLEMKF 914

RESULT 20
US-08-478-373-11
; Sequence 11, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:Vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-11

Query Match 4.5%; Score 180.5; DB 2; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

QY 5 TLKPI-----VLSILINTPLLAQH---ETEQSVG-----LETVTVVG-KSRP 44
DB 2 TKKPYRLSIICLLISCVKAEQSIKDTKEAISSEVDQTOSTDESELETISVTAEKIRD 61
QY 45 RATSGLLHTSTASDKII--SGDTLQKAVNLGDALDGVPGIHASQYGGGASAPV-IRGQT 101
DB 62 RQDNEV---TGLGKIITSEISREQVNLNLDTRYDPGISVVEQGGAGSSGYSIRGMD 117
QY 102 GRRIRKVLNHHGTGDMADSPDHAIWDTALSCQVEILRGPVTLI--YSSGNVAGLVDA 159
DB 118 RNRV-----ALLVDGLPQTQSVYVQSPVAVRSYGSGTGAINIEYE 158
QY 160 DKGIKPKMPENGVSGLGRLSSGNLEKLTSGINIGLG-KNFVLHTEGLYR----- 210
DB 159 NVKAVE-ISKGSSEYNGALAGSVTFQSKAADILEGDKSWGIGTKNAYSSKNKGFTH 217
QY 211 -----KSGDY---AVPRYRNLKRLPDSRRFANGQHR-AVLGWRKRFYRRYTSRRDQY 260
DB 218 SLAVAGKQSGFDGVALYTORN-----SIETQVHKDALAGVQS--YHRLIAKPEQ- 265
QY 261 GLPAHSHEVDDCHADIIWOKSL-INCRYLQLYPHLLTEEDVY-----DNP----- 305
DB 266 --SAYFMQDECPKPDYNSCLFFAKRPAILSSQRETVSVDYTGANRIKPNPMKYESQS 323
QY 306 -GLSCGFHDDDAHAHAHNGKPEWIDLRNKRYELRAEWKQPFPGFEALRVHLNRNDYHDE 364
DB 324 WFLRGYHFSQ---HYIGG--IFETQKQFDIR---DMTFAY-----LRSTE 364
QY 365 KAGDAVENFNNQTONA--RIELRHQPIGR-----LKGSGVGYLQKQSSA 408
DB 365 KRDDSSGSFPKQDYGAYQRIE-----DGRGVNYSGLYFDBHHRKQRVGIEYIYENKNK 419
QY 409 LSATSENV---KOPMLLNKVOH-----YSFFGVE----- 435
DB 420 AGIHDKAVLSANQOINILDSYMOHTHCSLYPNPSKNCRTREPKPYSYHSDRNVYKEKH 479
QY 436 -----QANW-----DNFT-----LEGGVYVEKQKASIRYDKALIDR 466
DB 480 MLQINLEKKIQNWLTHTQVFNLFQDDFTSALQHKDYLTERVTATAKSIS---EKANETR 536

```

```

QY 467 ENYIKO-----PLPDLG-----AHROTARSGFALSQNMWYFTPQHKLSL----- 503
DB 537 RNGYKKQPYLYPKPTVGVVQDHCDYKGNSSNVRDCKVRLIKGKNYYFAARNMALGKYV 596
QY 504 -----TASHQE-----RLPSTQEL 517
DB 597 DLGLGIRYDVSRTKANESTISVGKFNKFNWNTGIVIKPTWMLDLSYRLSTGFRNPSFAEM 656
QY 518 YAHGKHVATNTFEVGNKHLNKERSNNIELALGVEGDRWQVNLALYENRFQNYI-YAQTIN 576
DB 657 Y--GWRYGNNSEVYVYKPKPETSRAQERCLALKGDFGNIEISHFNSAYRNLIAPAEELN 714
QY 577 ---DGRPKSIEDSEMULVRYN-QSGADFYAGAEIYF----- 611
DB 715 KNGTGKANYGYHNAQNAKLGVNNTAQDLDFNGLWKRIPIYGVWATFAYNRVVKVQDKINAG 774
QY 612 -----KPTPRYRIGVSGD-----YVRGLKNLPSPGREDAYGNRPF 648
DB 775 LASVSYLFDALQPS-RYIIGLYGHPSTWNTGINTFTQSKAKSONELLGKR-ALGNN-- 830
QY 649 IAQDDQNAKPVPAARLGFHLKASLTDRIIDANLDYRVRFAONKLARYETETPPGHHMLNLA 708
DB 831 -SRDVKSTRKLTRA--WHI-----LDVSGYVYANK-----NIMLRIGI 865
QY 709 NYRNTRYGEWNNYVK-ADNLLN--QSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
DB 866 YNLFNRYVTWEAVROTAQAVNQHONQVGSYTRYAAS-----GRNYTLTLEMKF 914

```

```

RESULT 21
US-08-474-671-11
; Sequence 11, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973

```

```
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:Vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-474-671-11

Query Match          4.5%; Score 180.5; DB 3; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPLLAQAH---ETQSVG-----LETVTVG-KSRP 44
Db 2 TKKPYFRLSIISCLLISCVKRAETQSIKDKTEAISSEVDQTQSTEDSELETISVTKAKIRD 61
Qy 45 RATSGLLHTSADKII--SGDTLRQKAVNLGDALDGVPGIHASQYGGASAPV-IRGQT 101
Db 62 RKDNREV-----TCLGKIIITSISREQVLNIRDLTRYPDGISVVBQGRGASSYSIRGMD 117
Qy 102 GRRIKVLNHHGETGMADPSPDHAIMVDTALSQQVEILRGPVTL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPQTQSYVVQSLVARSYGSGTGAINIEIYE 158
Qy 160 DGKTIKPKMPENGVSGLRLSSGNLEKLTSGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKGSSEYNGALAGSVTFQSKSAADILEGDSWGTQTKNAYSSKNKGFTH 217
Qy 211 -----KSGDY---AVPYRNLRKLPSPRPANGQHR-AVLGWRKPRYRTYSRRDQY 260
Db 218 SLAVAGKGGDGVVAIYTORN-----SIETQVHKDALKGVQS--YHRLIAKPEDQ- 265
Qy 261 GLPAHSHEYDDCHADIIWQKSL-INKRYLQLYPHLLTEEDVDY-----DNP----- 305
Db 266 --SAFYVQDECPRPDDYNSCLPAKRPAILSSQRETVSDYTGANKRIKPNPKYESQS 323
Qy 306 -GLSCGFHDDDAHAHANGKPIDLRNKRVELRAEMKQPPFGEALRVHLNRNDYHDE 364
Db 324 WFLRGVHFSEQ---HYIGG-IFEFTQKFDIR--DMTFPAY-----LRSTE 364
Qy 365 KAGIAVENFFNQTQNA--RIELRHQPIGR-----LKGWGVQYLGQKSSA 408
Db 365 KRDDSSGSFYKQDYGVAYORIE-----DGRGVNYSAGLYFDEHKKQGVGIEYIENKNK 419
Qy 409 LSATSEAV-----KQPMLLDNKVQH-----YSPFGVE----- 435
Db 420 AGIIDKAVLSANQQNIILDSYMQHTHCSLYPNPSKNCRPTDKPYSYVHSRNVYKEKH 479
Qy 436 -----QANW-----DNPT-----LEGGVRVEKQKASIRYDKALIDR 466
Db 480 MLQNLLEKKIQNNLTHQIVFNGLGDFDFTSALOQKDYLTRVATATKSI--EKANETR 536
Qy 467 ENYIKQ-----PLPDLG-----AHRQTARSFALSGNWTPTPOHKLSL---- 503
Db 537 RNYGKQPYLPKTVGVVQVVDHCDYKGNSSNYRDKVRLIKGNKYVFAARNMALGKYV 596
Qy 504 -----TASHQE-----RLPSTQEL 517
Db 597 DLGLGIRYDVSRTKANESTISVGKFKNFSWMTGIVIKPTEWLDLSYRLSTGFRNPSAEM 656
Qy 518 YAHGKHVATNTPFVGNKHLNKRNNIELALGYGDRQWYNLRYRNFNGYI-YAQTIN 576
Db 657 Y--GWRVYGGNNSEYVGVKFKPETERNQBFGLAKGDFGNIIEISHFSNAYRNLIAPAEIN 714
Qy 577 ----DGRGPKSTEDDSEMKLVRYN--QSGADFYVAGSELYF----- 611
Db 715 XNGTGKANYGHNAQNAKLVGNVNIQAQLDFNGLWKRIPIYGHWATFAYNRVKVDQKINAG 774
Qy 612 -----KPTPRYIGVSGD-----YVRGLRKLNPSILPREDAYGNRPF 648

; 775 LASVSSYLFDALQPS-RYIIGLYDHPNSWTGINTMTFQSKAKSQNELLGKR-ALGNN-- 830
; 649 IAQDDQAPRVPAAELGPHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLC 708
; 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYTMANK-----NIMLRIGI 865
; 709 NVNRNTRYGEWNWYVK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGQVNVVKF 758
; 866 YNLFNRYVTWEAVRQTAQGVNQHVGSYTRYAAS-----GRNYTLLEMKF 914

RESULT 22
US-08-483-577A-11
; Sequence 11, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,577A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-577A-11

Query Match          4.5%; Score 180.5; DB 3; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPLLAQAH---ETQSVG-----LETVTVG-KSRP 44
```


Db 537 RNYKKQPYLPKPTVGVVQDHCYKGSNSYEDCKVRLIKGKYYFAARNMNLGKYY 596
Qy 504 -----TASHQE-----RIPSTOEL 517
Db 597 DLGLGIRYDVSRTKANESTISVGVKFNFSWNTGIVIKPTWELDLISYRLSTGFRNPSFAEM 656
Qy 518 YAHGKHVATNTFEVGNKHLNKRNSNIELALGYEGDRWQYNLALYRNFRNGYI-YAQTIN 576
Db 657 Y--GWRYGNNSEVYVGFPETSRNGQFGLAKGDFGNIHSHFSNAYRNLAFAEELN 714
Qy 577 ---DGRGPKSIEDDSEMKLVRYN--QSGADFYGAGEIYF----- 611
Db 715 KNGTGKANYGYHNAQNAKLVGNITAOIDFNLGKWRIPYGYWATFAYNRVVKDQKINAG 774
Qy 612 -----KTPRYRIGVSGD-----YVRGRLKPLSLPGREDAYGNRP 648
Db 775 LASVSSYLFDAIQFS--RIILGVDHPSNTWNTMTQSKAKSQNELKGR-ALGN-- 830
Qy 649 IAQDDONAPRVPAARLGFHLKASLDRIDANLDYRVFAQNKLYETRTPGHMLNLA 708
Db 831 -SRDVKSTRKTRA---WHI-----LDVSGYTMANK-----NIMLRGI 865
Qy 709 NYRNTRYGEMWYVYK-ADNLLN--QSVYAHSSFLSDTPQWGRSFTGCVNVKF 758
Db 866 YNLFNRYVYWEAVRQTAQAGAVNOHONGVSYTRYAAS-----GRNYTLTLEMKF 914

RESULT 25
US-08-649-518-11
; Sequence 11, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,518
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 914 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-649-518-11

Query Match 4.5%; Score 180.5; DB 4; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPELLAAH---ETEOSVG-----LETVTVVG-KSRP 44
Db 2 TKKPYFRLSIIICLLISCYVKAETOSIKDTKAISEVDTQSTEDSELETISVTAEKIRD 61
Qy 45 RATSGLLHTSTASDKII--SGDTLRQKAVNLGDALDGVPGIHAHQYGGGASAPV-IRGQT 101
Db 62 RKNQV---TGLGKIITSESRQVNLNIRDLTRYDFGISWVEQGRGASSGYSIRGMD 117
Qy 102 GRIKIVLNHGETGMADFPDPAHIMWDTALSOQVBEILRGPTVLL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPOTQSYVQSPVARSYSGTGAINEIYE 158
Qy 160 DGKIPEKMPENGVSGLRLSSNGLEKLTGGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKGSSEYNGALAGSVTFQSKAADILEDGKSWGIGTKNAYSNNKNGFTH 217
Qy 211 -----KSGDY---AVPRYENLRLPDSPPRFANGQHR-AVLGWRKRFYRTYSDBRDQY 260
Db 218 SLAVAGKQGGFGVAVYIQTNR-----SIETQVHKDALXGVQS--YHRLIAKPEDQ- 265
Qy 261 GLPAHSHEYDDCHADIIWQKSL--INKRYLQLYPHLLTEEDVQY-----DNP----- 305
Db 266 --SAYFMQDECPKDDYNSCLFFAKRPAILSSQRETVSVDYTGANRITKPNPMKYESQS 323
Qy 306 -GLSCGFPHDDDAHAHAHAKPWIDLRNKRYELRAEWKQFPFGEALRVHLNRNDYHDE 364
Db 324 WFLRGYHFEQ---HYIGG--IFEETQOKFDIR---DMTFPAY-----LRSTE 364
Qy 365 KAGDAVENFNNOTQNA--RIELRHQPIGR-----LKGSGVQVLGQKSSA 408
Db 365 KRDDSGSFYPKQDYGAYQRIE-----DGRGVNYSGLYFDEHHRKQRYGIEIYENK 419
Qy 409 LSATSEAV---KOPMLLDNKKVQH-----YSFGFVE----- 435
Db 420 AGIIDKAVLSANQONILDSYMOHTHCSLYPNBSKNCRPTDRKPYSYTHSDRNVYKEKN 479
Qy 436 -----QANW-----DNFT-----LEGGVVEKQKASIRYDKALIDR 466
Db 480 MLQINLEKTIQQNWLTHQIVNGLGDFDFTSALQHKDYLTERRVTATAKSIS--EKANETR 536
Qy 467 ENYVQK---PLPDLG-----AHRQTARSFALSNNWVETPQHKLISL--- 503
Db 537 RNYKKQPYLPKPTVGVVQDHCYKGSNSYEDCKVRLIKGKYYFAARNMNLGKYY 596
Qy 504 -----TASHQE-----RIPSTOEL 517
Db 597 DLGLGIRYDVSRTKANESTISVGVKFNFSWNTGIVIKPTWELDLISYRLSTGFRNPSFAEM 656
Qy 518 YAHGKHVATNTFEVGNKHLNKRNSNIELALGYEGDRWQYNLALYRNFRNGYI-YAQTIN 576
Db 657 Y--GWRYGNNSEVYVGFPETSRNGQFGLAKGDFGNIHSHFSNAYRNLAFAEELN 714
Qy 577 ---DGRGPKSIEDDSEMKLVRYN--QSGADFYGAGEIYF----- 611
Db 715 KNGTGKANYGYHNAQNAKLVGNITAOIDFNLGKWRIPYGYWATFAYNRVVKDQKINAG 774

Qy 612 -----KPTPRYRIGVSGD-----YRGLKMLPSLPGREDAYGNRPF 648
 Db 775 LASVSSYLFDIAQPS-RYIIIGLDHPSTWINTMTQSKAKSQNELKGR-ALGNN-- 830
 Qy 649 IAQDDQNAAPRVAARLGHKASLTDRIDANLDYRVRFAQKLAARYETRTTPGHMLNLA 708
 Db 831 -SRQVKSTRKLTTRA--WHI-----LDVSGYVMANK-----NIMLRGI 865
 Qy 709 NYRNRTRRGEWNNYK-ADNLLN--QSYVAHSSFLSDTPQGRSFTGGVNVKF 758
 Db 866 YNLFNRYVTWEAVRQTAQAVNQNVGSYTRYAAS-----GRNYTLTLEMKF 914

RESULT 26
 US-09-252-991A-30809
 ; Sequence 30809, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 30809
 ; LENGTH: 756
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-30809

Query Match 4.4%; Score 179; DB 4; Length 756;
 Best Local Similarity 20.7%; Pred. No. 4.5e-08;
 Matches 168; Conservative 104; Mismatches 266; Indels 272; Gaps 42;
 Qy 8 PIVLSILL---INTPLLAQAHTESQVLETTVVVGKSRPRATSGLLHTSTASDKIISGD 64
 Db 25 PLALSLSLFAFSAPALA-ADPVEQO-----MVVIGSRAPTRISELPGTVVWIER-EQLD 77
 Qy 65 TLROKAVNLGDALDG-VPGIHASQYGGGASAPVIRGQTRRIKVL---NHGEGDMAD 119
 Db 78 QQTQAGVPLKALGQLIPGLDICSQRTNNGQNL---GRSLVMDGVLSNRSRISQ 134
 Qy 120 FSPDHAIMVDTALSSQOQVEILRGVTLVLYSSGNVAGLVADGKIPEKMPENGVSQ----- 174
 Db 135 FD-----SIDPPNIEVWSG-ASAVYGGGATGGIINIVTK-----GVGGDTRFN 180
 Qy 175 -ELGLR-----LSSGNLEKLTSGGINIGLGNFVLHTEGLYKSGDYAV 217
 Db 181 TELGARGSQSHEDHLRAAQISGN--DLFNGRLAIYQKN-----GAAYDGSQDVL 233
 Qy 218 P-----RY-RNLKRLPSPRRFANGQRAVLGNRKRFRYRTYDRDRDQYGLPAHSHEY 269
 Db 234 TDITQTDQVNRSDVLMGSLGTFANG-HSLDILG----- 266
 Qy 270 DDCHADIIWQSLINKRYLYQLPHLLTEEDVDYD-NPGLSCGFHDDDDAHAAHNGKPI 328
 Db 267 -----LQYY-----DSGVDGDRGHDLDNRFDALGRAPYSIKGV 301
 Qy 329 DL-----RNKRYELRAEWKOPFPQGEALRYHLNRNDYHDEKAGDAVENFPNNTQONARIE 384
 Db 302 DLREPESKEHQFNATHPA-----EVLGHDLYLQAYYRNEK-----MAFNPPPT 346
 Qy 385 LRHQPIRLKSGVQVLOKSSALSATSEAVKQPMLLNKKVQHSFFVEQA-----NWDN 441
 Db 347 IRYSNTGAI--NYGTSY-----YSASQ-----QDTDYGMKALVKTWER 384
 Qy 442 FTLEGGVRVBEKQAS-----IRYDKALIDRENYVQKLPDLGAHQRTA 484

Db 385 ASLTGVGLDREKFTSDQMLFNLPLAASGGGLVASEQAKLGR-----YPDIDT---DS 434
 Qy 485 RSFALSGNWFYTPQHKSLTASHOERLPST-----QELYAHG-----K 522
 Db 435 RAFFLOGSKATDD--LTLASGVRQSQMSTDVSDFVAANOQILLIANGLGKTADAVPGGSK 492
 Qy 523 HVAITNTFEVGNKH-LMKERSNNIELALGYE--GDRWQYNLALRYRNFNYIY-----AOTL 575
 Db 493 DIDVNLVNVGAIYKMLNQOVWANYSEGPELDPKAYYGFGRYGAADGNHYPPLQGVSV 552
 Qy 576 NDGR-----GPKSIEDDSEMKLYRYNSGADFFGA 605
 Db 553 NDSPLDGIKTKQVELGWRHTDQALDTQAAFYSWSDKSIKYDSKTLAVLQOQTKRNYGL 612
 Qy 606 EGEIYEKPTPRYRIGVSGDYVRGLKNLPSLGR---EDAYGNRP-----FIA-QDDQNA 656
 Db 613 EGQATYWLDDHWQVGVNGLAIRSQEK-----VDGRWLKQDVTSSASPSKAGAFVGMKDDQRS 668
 Qy 657 PRVPAARLGHKASLTDRIDANLDYRVF 686
 Db 669 LRQQGV-----TFNLNDEFGNKIDGYALF 693

RESULT 27
 US-09-328-352-5942
 ; Sequence 5942, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 5942
 ; LENGTH: 761
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-5942

Query Match 4.3%; Score 174; DB 4; Length 761;
 Best Local Similarity 19.9%; Pred. No. 1.4e-07;
 Matches 180; Conservative 119; Mismatches 299; Indels 308; Gaps 46;
 Qy 7 KPIVLSIL---LINTPLLAQAHTESQ-----VGLTIVTVYVKSHPRTATSL 50
 Db 10 KRIQSVLSVSLASMMMAFAAQNEQDAEKTLEKPAEPVKLETIFVTAEQVKOSLGV 69
 Qy 51 LHTSTASDKIISGDTLRQKAV--NLGDALDGVPGIHASQYGGGASAPVIRGQ----- 101
 Db 70 -----SVITQEDLEKLPVRNDISDYVRMPGVNLT---GNSATGQGNRQIDIRG 117
 Qy 102 -----GRIKVLN--HGETGDM-----ADSPDHAIMVDTALSSQVEILRGV 143
 Db 118 MGPENTLILVDGKPIINSRNSRVYRKGWGERDTRGDSNWPABAI-----ESIEVLRGPA 170
 Qy 144 TLYSSGNVAGLVADGKIPEK-----MPEN---GVSGELGLRLSS----- 182
 Db 171 AARYGSAAGGVNIIITKVTNETHGSEVFTSQPEDSKESGNSRVGNFVSGPLIKDVL 230
 Qy 183 -----GNLEKLTSGGINI--GLGNFVLHTEGLYRK--SG-----DYAVPRY 220
 Db 231 YRLYNTNKTAEADDVDINKSIGST-AAGREGVKNKDISGLRLAWQATDQQTLLADISSSQ 289
 Qy 221 RNLEKLPDSRRFRANGQRAVLGNRKRFEYRPTYSRDRDOYGLPAHSHEDDCHADIIWQ 280
 Db 290 GNI---YSGDSQLNANAEADAILS--QLIGKETMTNRDYSAL---THE-----GDSWKG 337
 Qy 281 SLINKRYLYQLPHLLTEEDVDYDNPGLSCGFHD-----DDAHAAHNGKPIWDLNRKY 335
 Db 338 SKLVAQYDKTHNKELPE-----GLAGSVEKINNLLDDKATS-----SLETL 378


```

QY 9 IVLSTILLINTPLLAQH--ETBQSVGLTETVTVGKSRPRATSGLLHSTASDKIISGDTL 66
Db 28 VTLSTILLASPSALARENPPAEPALVVTGKRRS-----LKDTSVSLTSARDI 83
QY 67 ROKA---VNLGALDQVGIHASQYGGGASAPVIRGQTR-----RIK 106
Db 84 DRKQTNASVAEVINGSPV---VYDSVGAPIIRGQDTQGPENNGQNVFVGTVPRATIN 140
QY 107 VLNHHGETGDMADFPDHAIMVDTALSOQVEILRGPVTLVSSGNVAGLVDAQKIPEK 166
Db 141 LOGHYLNINEMF--FGATSVMDVDS-----IEVFRGPOTTSQANAIAGAI--IVNTKOPTF 193
QY 167 MBENGVSGLGLRLS-----SGMLEKLTSGGINIG--LGNFVHLTEGLYRKSGDYAVP 218
Db 194 SPEAYQAEIGSYHSRRSSIAVSGPLAQDFAGELAVDVAGRTTFI-----DYDNP 243
QY 219 RYRNKLRLPDSPPRFANGQRAVLGWRKRYFRRTYSDRRDQVGLPAHSHEVDDCHADILW 278
Db 244 KFO-----DSG---TDQDFRAL-----NARAKLLW 265
QY 279 QKSLINKRYLQVPHLLTEEDVDYDNPGLSCGF---HDDDDAHAAHNGKFWIDLNRKRY 335
Db 266 LPSGI-----PGLSEKFTSHDNRPTQOEAATRF--DKLDHRS 303
QY 336 ELRAEWKQ-----PPFGEALRVHLNRNDY-----HDEKA---GDAVENFNNQT 378
Db 304 TTMPSEWQDNTSILDVAYDLNDGIRL--FNQAYSLSVHRTTGAGGEGDA--DIRQKNAS 361
QY 379 QNARIELRHOPIGLRKGSGVQVLOGKSSALSATSEAVKQPM---LDNKKVQHYSPFGVEQ 436
Db 362 NESRVFGEQE--DRISGMGVY-----ARTDETILHRLGLSAPDDTKNGLVFG--E 412
QY 437 ANW---DNFTLEGVAVEKQKASIRVDKAL-----IDRENYKQPLPDIGAHRQTARSPAL 489
Db 413 LNYRLSDRWTLTGLRYQEDRIERSGNSVLAPEPLDYQKTFGAFLE-----KVSIAFAA 466
QY 490 SGWNYTPQKLSHTASHQERLPSTQELVAGHGVATNTEFVGNKHLNKRSENNIELALG 549
Db 467 TPDW-----TVGGLVSRG-----YNGGVSLNLTFRN---WAYF 497
QY 550 YEGDRWOYNI---ALYRNRF---GNYIYAQTLNDGRGPKSIEDDSEMKLVRYNOSGADF 602
Db 498 KEETINWYELFTRASLLDGLRLNGLNFEMDFKDAQYINPVVVSQVQAQSVTINAEKHA 557
QY 603 YGABEGLYFPTRYIGVSGDVVGRGLKNLPSLPGREDAYGNRPFIADQDQNAQRPVAA 662
Db 558 YGMELDDYRLRDLRLKASAGVIRTRIDEMSGNTG-----YEHNEF-----ARSPGY 605
QY 663 RLGFHLKASITDRIDAN-----LD--YRVFPAQNKUARYETRTPGHHMLNLGANYRNTR- 715
Db 606 TLSFGPSWDTVERLNLNAQVRYLDGYYSDTANTKAYSIKAYT---LTDARASYRFNDQV 661
QY 716 --XGEMWNVYK 724
Db 662 QLYG-----YVK 668

```

RESULT 30

```

US-09-252-991A-25815
; Sequence 25815, Application us/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 25815
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25815

Query Match 4.2%; Score 169; DB 4; Length 595;
Best Local Similarity 20.8%; Pred. No. 2.8e-07;
Matches 159; Conservative 97; Mismatches 250; Indels 258; Gaps 42;

QY 78 DGVGFIHASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADFPDHAIMVDTALSOQVE 137
Db 5 DGIP-----ASTPDGQCG-----AATLNDVAERIE 30
QY 138 ILRGPTVLLYSSGNVAGLVDPV--ADGKIPEKMPENGVSSELGLRLSSGNLEKLTSGGINI 195
Db 31 VLKGPASTIYGS--NAGVTOMFSDGQAPRVGAATVGSQG--LSRNHL--YTEGE--- 82
QY 196 GLGKNEV---LHTEGLYRKSGDYAVPR---YRNKLRLPDSPPRFANGQRAVLGWRK 246
Db 83 GDGVGFLVDASRMDTDG--YR--DHSAAARDQTFAKLNPRPDADSLA----- 126
QY 247 RFYRRYS-----DRRDQYGLPAHSHEYD-----DCHADIIWQKSLIN--KRY- 287
Db 127 ---LIYSLEQNDTDEPLGQTDWYKHDPRSVTANAEIYDTRKSIDHQAGNTERIFG 182
QY 288 ---LQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKFWIDLNRKRYELRAEWKQ 344
Db 183 EATLQVNAVYVGRSVVQYQAIKQIGC-----ESNPRCQRNGAVIDFDRDFHGGTVRMLQP 238
QY 345 F---RQFEALRVHLNRNDYHDEKADCAVENEFNNTQONARIELRHOPIGLRKGSGVQY 401
Db 239 VSOAPGELNLTVGLD--YDQSRDDRG--YQN--FNGDGLGVKGLRRDEVD----- 284
QY 402 LGQKSSALSATSEAVKQPMLLDNKVKQHYSGFFGVEQANW---DNFTLEGVRYVEKQKASIRY 459
Db 285 ---TATS---LDPVL-----QASWAIDAWTLQAGVYRHSMTKMEV-- 317
QY 460 DKALIDRENYKQPLPDIGAHRQTARSPALSGNWFTTPOKLSLTASHOERLPSTQEL--Y 518
Db 318 ---DDRYLSNGSDASGRYRKNTPSFVYVYAFITPDHGLYLSAGKGFETPTQAEWAY 370
QY 519 AHGKHVATNTEFVGNKHLNKRSENNIELALGYEGEDRW---QYNLALYRNRFNGYIYAQTL 575
Db 371 A---PVAANAPDVFNGLKPAISSQVE--AGLKAHLGNTRVNAALFOIRTEDEIVVASS 425
QY 576 NDGRGPKSIEDDSEMKLVRYNOSGADF-----YGABEGLI-----YFKEPTPYRIGV 621
Db 426 LGGR-----TSYQNAAGKTLRREGFELGLESELSEHWNANLAYTRLSATYDSDF 472
QY 622 -SGDYVRGRGLKNLPSLP-----GREDAYGNRPFIADQDQNAQRPVAA 662
Db 473 EAGGKTIGKGLHPGVPESSLFGEVLVWKPAGISMGWEGYRSQYV--EDSNSEKAAPS 530
QY 663 RLGFHLKASITDRIDANLDYRVFPAQNKUARYETRTPGHHMLNLGANYRNTRYGEWNY 722
Db 531 ---YAVF--NMRTRFE-----QRLGTWAFH 550
QY 723 --VKADNLLNQSVAHSSFLSDTPOM-----GRSFTGVGVVVKF 758
Db 551 QLVLDNLNLFDRQ--YGVSVIVGDGNRRYEAAPGLSWYAGVGEY 593

```

RESULT 31

```

US-09-252-991A-30935
; Sequence 30935, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

Query Match	4.1%;	Score 164.5;	DB 4;	Length 643;
Best local Similarity	19.0%;	Pred. No. 8.6e-07;		
Matches 145;	Conservative 105;	Mismatches 288;	Indels 225;	Gaps 36;
QY	24	AHETEQSGLGVTVV-VGKSRPRATSGLLHSTPASDKIIISGPTLRQKAVNLGDALDGVPG	82	
Db	49	AEDATDQTDLPVITASKSAEKASVAPARISIIIEPKIVEQSPIAE----	104	LPHLMSDAA
QY	83	IHASQYV--GGASAPVIRGQTGRKIVLNHGHGTGMADFSPDHAIWDTALTSQOQEILR	140	
Db	105	INMQSGGLGQTSTIFIRGTNSSEHAIL--RDGARLNTASTGAANLAFIDTTIKQIEILK	163	
QY	141	GPVILLYSSGNVAGLVADVADKPIKEMPENGVSQELGRLSSGNLEKLTSGGINCLGNK	200	
Db	164	GPASVLYGTDAIGGVQII--SKTPEK--TSAFVTGEIG-----ENKTKSI--VGAD	210	
QY	201	FVLHTEGLYKRSQDYAVPRVRLNKLPDPSRPFANGCHRAVLGWRKFRFRYYSDRRQY	260	
Db	211	LA-----ENGFYAQVQRQLE-----SDG-----SRITDLK	236	
QY	261	GLPASHSEYDCHADIIWQSLINKRYLIQYLPHLLTEED---VDY--DNPGLS--CGFHD	313	
Db	237	GNDIKKASYD-----QKGFSKVGVE-----KEDFGASVDYQTNEGTSQYDFTFY	281	
QY	314	DDDAHAHAHGKPHWIDLENKRYELRABWKQPPFGFALRWLN-----RNDYHDEKA	366	
Db	282	DGSLTSQ-----DFKNELLNIRG-----RVNLNSDTSLNARLSQFKDELD	321	
QY	367	GAIVENFNNQTONARIELARHQPICRLKGSWGQVYLQCKSSALSASEAVKQPML----	421	
Db	322	QNGSDFVHSITTKAEV-----YGKW--QFTSSQVNLAGVTHQINDGDVLSYSGP	369	
QY	422	LNKVOHYSFPGVEQANWNTLEGGVRVEKQKASTRYDKALLIDRENYKQPLFDLGAHR	481	
Db	370	YDEDVNSTGYFVQHQYQNNGLNTQGVRRV-----DNEKY-----GTH--	407	
QY	482	QTARSFALSGNWYPTPOHKLSLTASHQERLPSTOELYAHGKHVATNTFEVGNKHLNKERS	541	
Db	408	-TVAQGAI--RYQLLPLTISIYANIGSFAKPTLNDMYGSG-----GNENLKPEES	454	
QY	542	NNIELALGYEGDRWQYN-----LALYRNRFGNYIYAOTLN--DQGGPKGSTEDDSEMKLVRY	595	
Db	455	ISYEYGI---DQKLNYNISGLSAYYTKIDNLIESRCIAVCGNDGWINTFPVQYNINIDRA	511	
QY	596	NQSGADFDYV--ABGEIYFKPTEPRYIGVSGDYVRGLKNLPLCPGREDAYGNRPFFIAQD	653	
Db	512	SNRGEVYANWRDDLFYKSSVYVKAIND-----ETD	544	
QY	654	QNAVRVPAARLGFHLKASLTDRIDANLDYY-----RVFAQWKLARYETRRTPGH-----	701	
Db	545	QELSSRP--ROSFTVSTGLQW-----EHYGLSISLSAKSKADYQDQFTPGYTVVDNFGY	596	


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,435
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-462 MIS.vg
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-435-5

Query Match 4.0%; Score 163; DB 2; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

Qy 5 TLKPI-----VLSILLINTPLAQAH---ETEQSVG-----LETVTWG-KSRP 44
Db 2 TKRPYRLSIICLLISCYVKAETQSIKDTKEAISSEVDTQSTEDSELETISVTAEKVRD 61
Qy 45 RATSGLIHTSTASDKII--SGDTLRQAVNLGDALGVFGIHASQYGGGASAPV-IRGT 101
Db 62 RKNEV---TGLGKIKTSESISREQVNLIRDLTRYDPGISVVEQGRGASSGYSIRGWD 117
Qy 102 GRIKVL-----NHCETGDMADFPDHALIMVDTALSQQVEILRGP 142
Db 118 RNRVALLVDGLPQTQSVVQSPVLVARGSGYSGTGAINIEYENV-----KAVISKGG 169
Qy 143 VTLIYSSGNVAGLV-----DVADG-----KIPEKMPENGVSGLGLRLSSGNLE 186
Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTQKNAYSNKNGKPTHS LAVAGKGGFE 229
Qy 187 KLTSGGINIGLGN-----FVLHTE--GLYKRSGLYAVPRNLKR 225
Db 230 -----GVAIYTHRNSIETQVHKDALKGVQSYDRFIATTDQSAYFVMQDECLDGYDKCT 284
Qy 226 LPDSRFRFANGQRAVL-----GWRKRFRYRTYSRRDQY--GLPA 264
Db 285 SPKPEATLSQRETVSDYDGTANIKPNPKMYESQSW---FLRGYHFESEQHYIGGI-- 339
Qy 265 HSHYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHNG 324
Db 340 --FEFTQCKFDI-----RDMTFPAYLEPTEDKDLQ-----SRFPYKQDYGAVQHIG 384

```



```

; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-337-483-5

Query Match 4.0%; Score 163; DB 2; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

QY 5 TLKPI---VLSILLINTPLIAQAH---ETEQSVG-----LETVIIVG-KSRP 44
Db 2 TKKPYFRLSIISCLLSYCVKAEQTSIKDIKEALSEVDFTQSTEDSELETISVTAEKVRD 61
QY 45 RATSGLHTSTASKII--SGDTLRQKAVNLGALDGVPGIHASQYGGASAPV--IRGT 101
Db 62 RKNDEV---TGLGKIITSEISIREQVLRNIRLTRYDPCISVVEQGRGASSGYSIRGMD 117
QY 102 GRRIKVL-----NHHGETGDMADSPDHAIMVDTALSQOVELRGP 142
Db 118 RNRVALLVDGLPQTSYVQSVPLVARSYSGTGAINEIENV-----KAVEISKGG 169
QY 143 VTLISYSGNVGLV-----DVADG-----KIPEKPNVSGVSGELGLRLSGNLE 186
Db 170 SSSEYGNALAGSVTFQSKSAADILEGDKSWGIGTQNAYSKKNKGFTSLAVAGKQGGPE 229
QY 187 KLTSGGINIGLGN-----FVLHTE--GLYKSGDYAVPRYRNLR 225
Db 230 -----GVAIYTHRSIETQVHKDALKGVQSVDRFIATTEQSAFYVQWDECLDGYDKCT 284
QY 226 LPDSPRRPANCQRAVL-----GWRKRFYRTYSRRDQY--GLPA 264
Db 285 SPKRPATLSTQRETQSVSDYTGANRIKPNPMKYESQSW---FLRGYVHFSEQHYIGGI-- 339
QY 265 HSHEVDCHADIIWQKSLINKRYLQLYPHLLTDEVDYDNPGLSCGFPHDDDDAHAAHNG 324
Db 340 --PFTQKQFQI-----RDMTFPAYLRPTEDKDLQ-----SRFPYKQDYGVQCHIG 384
QY 325 KPWIDLNRKRYELRAEWKQPPGFPEALRVHLNRNDYHDEKAGDAVENFFNQTVARIE 384
Db 385 ----DGRGVKY-----ASGLYFDEHHRKORVG--IHYIYENKNAKAGIID 422
QY 385 LRHQPIGRKLSGWSGVYLGQKSSALSATSEAVKQPMILLDNKVOH-----428
Db 423 -----KAVLSANQQNIILDSVMRHTHCSLYPNPSKNCRPT 457
QY 429 ---YSFPGVE-----QANW-----DNFT-----443
Db 458 LDKPYSYTHSRNRYVYKHEKNNMLQNLLEKTIQQNWLTHQIAFNIGFDFTSALQHDYLTR 517
QY 444 --LEGGVVEKOKASIRYDKALIDRENYKQPLP-----DLCAHROTARSA-----488
Db 518 RVIATASSISEKRGAEARNGL---QSSPYLYPTPKAELVGGDLNLYQKSSNYSDDCKVRL 574
QY 489 LSG-NWYTPQHKULSLTASHQERLPSTQELYAHOKHVAATNTEVGNKH--LNKRSNNIEL 546
Db 575 INKKNYFAARNMAL-----GKYV---DLGLGMEYDVSRTKANESTI 614
QY 547 ALG-YEGRWQVNLALYRNRFNGNYIAOTLNDG-RGPKSIEDDSEMKLVYNQSGADFYG 604
Db 615 SVGKEKFNFSWNTGVIKTEWLDLSY--RLSTGRNPSF-----AEMYGWYGGKDTIYI 668
QY 605 AGEIYFKP-----TPRYIGVSGDYVRGRKLNLSLPGREDAYGNRPFIAQDQONAPRVP 660
Db 669 GK-----FKPETSIRNOEFGALKGDFGNTIEISHF-----SNAYRNLIATAEELSCKGTG 718

```

```

QY 661 AARLGFH-----LKASLTDRIDAN-----LDYRVVFAQNKLA-----692
Db 719 KENYGVHNAQNAKLVGVNITAOGLDFRGLMKRIPYGYIATFAYNRVVKDQKINAGLASVS 778
QY 693 --RYETRTPGHMLNLGANYRRNTRYGEMWVYKADNLLNQSVVAHSSFLSDTPQMGRSF 750
Db 779 SYLFDATQPSRYIIGLYDHPSTN-----WGIK-----TMFTQSKAKSQNELLGKRA 825
QY 751 TG--GVNVK 757
Db 826 LGNNSRNVK 834

RESULT 36
US-08-478-373-5
; Sequence 5, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-5

```

Query Match 4.0%; Score 163; DB 2; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;


```

187 QY 187 KLTSGGINIGLGN-----FVLHTE--GLYKSGDYAVPRYNLKR 225
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 Db -----GVAIYTHRSIETQVHKDALKGVQSYDRFIATTEQSAFYVMQDECLDGYDKCKT 284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
226 QY 226 LPDSPRFANGQHRVAVL-----GWRKRFRVRRYTSRRDOY--GLPA 264
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
285 Db 285 SPKRPAULSTQRETVSVDYTGANRIKPNPMKYESQSW---FLRGYHFSBOHYIGGI-- 339
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 QY 265 HSHEYDDCHADI IWQSLINKRYLQYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNG 324
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
340 Db 340 --FEFTQCKFDI-----RDMTFPAYLRPTEDKDLQ-----SRPFYKQDYGAYQHIG 384
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325 QY 325 KPWIDLRNRYELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNQTONARIE 384
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
385 Db 385 ---DGRGVKY-----ASGLYFDEHHRKORVG--IEYIYENKNKAGIID 422
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
385 QY 385 LRHOPF1GLKSGWGVQYLGOKSSALSATSBAVKQPMLLDNKVQH----- 428
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
423 Db 423 -----KAVLSANQQNIILDSYMRHTHCSLYPNPSKNCRPT 457
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
429 QY 429 ---YSFEGVE-----QANW-----DNFT----- 443
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
458 Db 458 LDKPYSYHSDRNRYKKEHNMQLNLEKKIQNWLTQIAFNGLFDDFTSALQHKDYLTR 517
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
444 QY 444 --LEGGVREKOKASIRYDKALIDRENYKQPLP-----DLGAHROTARSFA----- 488
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
518 Db 518 RVIATASSISEKRGEARNGL---QSSPYLYPTPKAELVGGDLCTNYGKSSNYSDCKVRL 574
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
489 QY 489 LSG-NWVETPQKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKH-LNKERSNIEL 546
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
575 Db 575 IKGNNYFAARNMAL-----GKYV-----DLGLGMYDVSRYKANEI 614
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
547 QY 547 ALG-YEGDRWQYNIALYRNFGNYIYAQTLDNG-RGPKSIDDSSEMKLVRYNOSGADPYG 604
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
615 Db 615 SVGRKFNSWNTGVIKPTWELDSY--RLSTGPRNPSF-----AENYGMRYGCKTDVYI 668
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
605 QY 605 ABGEIYFKP---TPRYRIGVSGDYVRGLKNLPSLPGREDAGNRPFIADQDNAPRVP 660
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
669 Db 669 GK---FPETSRNQEFGLAKGDFGNIEISHF-----SNAYNLIAPABELSKNGTTG 718
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
661 QY 661 AARLGFH-----LKASLTDRIDAN-----LDYRVFAONKLA----- 692
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
719 Db 719 KGNVYHNAQNAKLGVGNVITAGLDFNGILWKRIPIYGMVATFAYNRVVKQDKINAGLASVS 778
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
693 QY 693 --RYETRPFGHMLNIGANRYNTRYGEMWVYKADNLLNQSVYAHSSPLSDTFQWGRSP 750
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
779 Db 779 SYLFDALQPSYIIGLYDHFST-----WGIK-----TMTQSKAKSQNELIGKRA 825
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
751 QY 751 TG--GVNVK 757
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
826 Db 826 LGNNSRNK 834
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 38

US-08-483-577A-5

Sequence 5, Application US/08483577A

Patent No. 6015688

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena

APPLICANT: Harkness, Robin

APPLICANT: Schryvers, Anthony

APPLICANT: Chong, Pele

APPLICANT: Gray-Owen, Scott

APPLICANT: Yang, Yan-ping

APPLICANT: Murdin, Andrew

APPLICANT: Klein, Michel

TITLE OF INVENTION: Transferrin Receptor Genes

NUMBER OF SEQUENCES: 160

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

```

COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,577A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-511
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-577A-5

```

Query Match 4.0%; Score 163; DB 3; Length 913;

Best Local Similarity 17.5%; Pred. No. 2.2e-06;

Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

```

QY 5 TLKPI-----VLISILINTPLLAQAH-----FTEQSVG-----LEFTVTVG-KSRP 44
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 TKKPYFRLLISCLLSICVKAETQSIKDKTEALISSEVDQTQSTEDSELETISVTAEKVRD 61
QY 45 RATSGLLHTSTASDKII--SGDTLRQKAVNLGALDGVGIGHASOYGGGASAPV-IRGQT 101
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 RDNREV-----TGLGKIYTSISREQVLIARDLTRYDPLGVVVEQGRGASSYSIGMD 117
QY 102 GRRIKVL-----NHHGETGMADFSPDHAIMVDYTSALSOQVEILRGP 142
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 RNRVALLVDGLPQTQSVVQSPVARSQSGYSGTGAINIEYENV-----KAVEISKGG 169
QY 143 VTLLYSYSGNVAGLV-----DVADG-----KIPKMPENGVSGELGLRLSSGNLE 186
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 SSSEYGNALAGSVTFQSKNAADILEGDKSWGQTKNAYSNNKNGFTSHLAVAKQGGFE 229
QY 187 KLTSGGINIGLGN-----FVLHTE--GLYKSGDYAVPRYNLKR 225
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 -----GVAIYTHRSIETQVHKDALKGVQSYDRFIATTEQSAFYVMQDECLDGYDKCKT 284
QY 226 LPDSPRFANGQHRVAVL-----GWRKRFRVRRYTSRRDOY--GLPA-264
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
285 SPKRPAULSTQRETVSVDYTGANRIKPNPMKYESQSW---FLRGYHFSBOHYIGGI-- 339
QY 265 HSHEYDDCHADI IWQSLINKRYLQYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNG 324
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
340 --FEFTQCKFDI-----RDMTFPAYLRPTEDKDLQ-----SRPFYKQDYGAYQHIG 384
QY 325 KPWIDLRNRYELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNQTONARIE 384
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
385 ---DGRGVKY-----ASGLYFDEHHRKORVG--IEYIYENKNKAGIID 422
QY 385 LRHOPF1GLKSGWGVQYLGOKSSALSATSBAVKQPMLLDNKVQH----- 428

```

```

Db 423 -----KAVLSANQONILDSYMRHCSLYPNPNSKNCRPT 457
Qy 429 -----YSFGVE-----QANW-----DNFT-----443
Db 458 LDKPSYSHSDRNVYKEKHNMQLNLEKKIQONMLTHQIAFNGLGDFDFTSALQHKDYLTR 517
Qy 444 --LEGVRVKKASIRYDKALIDRENYKQPLP-----DLGAHQRTARSA-----488
Db 518 RVATASISSEKRGARNGL---OSSPYLPTKAEVLGGBLCLNYGKSSNYSDDCKVRL 574
Qy 489 LSG-NWYFTPOHKLSTASHOERLPSTQELVAHGKHVATNTFEVGNKH-LNKERSNNIEL 546
Db 575 IKGNVYFAARNMAL-----GKYV-----DLGLGMYDVSRTKANESTI 614
Qy 547 ALG-YEGDRWQYNLALYRFGNYIYAQTLDNG-RGPKSIEDDSEMKLVYQNSGDFYQ 604
Db 615 SVGKFKNSWNTGVIKTEMLDLSY--RLSTGPRNFSF---AEMYGWRYGKDTUUYI 668
Qy 605 ARGEYFKP---TPRYRIGVSGDYVRGRLKNLPSLEGREDAYGNRRFFIAQDDQNAPRVP 660
Db 669 GK---FKPETSQNEFGLALKGDFGNTIEISHP-----SNAYRNLIATAFAEELSKNCTTG 718
Qy 661 AARLGFH-----LKASLTORIDAN-----LDYRVFPAOKLA-----692
Db 719 KGNVYHNAQNAKLVGVNITAQDLFNGLWKRIPIYCGWATFAYNRVYKVDQKINAGLASVS 778
Qy 693 --RYETRTPGHMLNGLANRYRNTRYGEWNYVAKADNLLNQSVYAHSSFLSDTTPQMGRSF 750
Db 779 SYLFDALQPSRYIIGLYDHPST-----NGIK-----TMTQSKAKSQNELGKRA 825
Qy 751 TG-GVNUK 757
Db 826 LGNRSNRVK 834

```

RESULT 39

```

US-08-897-438-5
; Sequence 5, Application US/08897438
; Patent No. 6262016
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,438
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-897-438-5

Query Match          4.0%; Score 163; DB 3; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

Qy 5 TLKPI---VLSILLINTPLLAQH---ETQSVG-----LETVVVVG-KSRP 44
Db 2 TKKPYFRLSIISCLISCVKAEQTQSIKDKTEAISSEVDYDTQSTEDSELETISVTAERVD 61
Qy 45 RATSGLLHTSASDKII--SGDTLRQKAVNLGDALDGVPGIHASQYGGGASAPV-IRGQT 101
Db 62 RKNVEV---TGLKLIKTSISISREQLNTRDTRDPGLISVVEQGRGASSGYSIRGMD 117
Qy 102 GERIKV-----NHHGTGMDADPSPDPAHIMVPTALSSQQVEILRGP 142
Db 118 RNRVALLVDGLPQTQSVYVQSPLVARSYSGYSGTAINETIYENV-----KAVEISKGG 169
Qy 143 VTLLYSSGNVAGLV-----DYADG-----KIPEKMPENGVSSELGLRLSSGNLE 186
Db 170 SSSEYNGGALAGSVTFTQSKSAADILEGDKSWGIOQTKNAYSKNGKGFTHSLAVAKQGGFE 229
Qy 187 KLTSGGINIGLKN-----FVLHTE--GLYRKSGDYAVPRYRLMKR 225
Db 230 ----GVAIYTHRSIETQVHKDALKGVSYDRFTATTEDQSAVFVMDCLDGVKCKT 284
Qy 226 LPDSRFRFANGQRAVL-----GHRKEFYRTYSDRDQY--GLPA 264
Db 285 SPKRPTLSTQRETQSVSDYTGANRIKPNPKMYESQSW---FLRGYHFSQHYIGGI-- 339
Qy 265 HSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNFGLSCGPHDDDDAHAAHNG 324
Db 340 --FEFTQOKFDI-----RDMTFPAYLRPTEDKDLQ---SRPFYPKQDYGAYQHIG 384
Qy 325 KPWIDLNRKRYELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNQTCNARIE 384
Db 385 ----DGRGVKY-----ASGLYFDEHHRKQRVG--IEVIYENKKNAGIID 422
Qy 385 LRHQPIGRLLKGSWGQVYLGQSKSALSATSSEAVKQPMLLDNKVQH-----428
Db 423 -----KAVLSANQONILDSYMRHCSLYPNPNSKNCRPT 457
Qy 429 ----YSPFGVE-----QANW-----DNFT-----443
Db 458 LDKPSYSHSDRNVYKEKHNMQLNLEKKIQONMLTHQIAFNGLGDFDFTSALQHKDYLTR 517
Qy 444 --LEGVRVKKASIRYDKALIDRENYKQPLP-----DLGAHQRTARSA-----488
Db 518 RVATASISSEKRGARNGL---OSSPYLPTKAEVLGGBLCLNYGKSSNYSDDCKVRL 574
Qy 489 LSG-NWYFTPOHKLSTASHOERLPSTQELVAHGKHVATNTFEVGNKH-LNKERSNNIEL 546
Db 575 IKGNVYFAARNMAL-----GKYV-----DLGLGMYDVSRTKANESTI 614
Qy 547 ALG-YEGDRWQYNLALYRFGNYIYAQTLDNG-RGPKSIEDDSEMKLVYQNSGDFYQ 604

```

```

615 SVGKFNFSWNTGIVIKPTWLDLSY--RLSTGFNPSF---ARMYGMRYGKDTDVI 668
QY
605 ABGEIYFKP---TPRYRIGVSDYVRGLKNLPSLPGREDAYGNRPFLAODDQNA PRVP 660
Db
669 GK-----FKPETSNGQFGLKGDGFGNIBSHF-----SNAYRNLIAPABELSKNGTIG 718
QY
661 AARLGPH-----LKASLTDRIDAN-----LDYRVFAONKLA----- 692
Db
719 KGNVGHNAQNAKLVGNVITAOQLDFNGLWKRIPIYGYATFAYNRVKVDQKINAGLASVS 778
QY
693 --RYETRTPGHMLNIGANVRNTRYGEWNWYVKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db
779 SYLFDALQPSRYIIGLYDHPSTNT-----WGIK-----TMTQSKAKSONELLGKRA 825
QY
751 TG--GVNVK 757
Db
826 LGNNSRNK 834

RESULT 40
US-08-637-654-5
; Sequence 5, Application US/08637654
; Patent No. 6358727
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Hartness, Robin E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew D
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,654
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/CA94/00616
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-595
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-637-654-5

```

Query Match 4.0%; Score 163; DB 4; Length 913;
 Best Local Similarity 17.5%; Pred. No. 2.2e-06;
 Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

```

QY
5 TLKPI-----VLSILLINTPLLAQAH---ETBOSVG-----LETVTWVG-KSRP 44
Db
2 TKKPYFRLSIISCLILSCYVKAETOSIKDKAEISSEVDSTQSTEDSELETISVTAEKVRD 61
QY
45 RATSGLLHSTASDKII--SGDTLRQKAVNLGDALDVGPGIHAISOYGGGASAPV-IRGT 101
Db
62 RKNFV-----TGLGKIITKTSISIRSEQVNIIRDLTRYDPGISVVEQGRGASSYIRGMD 117
QY
102 GRRIKVL-----NHHGETGDMADFPDPAIMVDTALSOQVILRGP 142
Db
118 RNRVALLDCLPQTQSVYVQSPVARSYSGTGCAINEIYENV-----KAVEISKGG 169
QY
143 VTLLYSSGNVAGLV-----DVADG-----KIPEKPEENGVSCELGLRLSSGNLE 186
Db
170 SSSEYNGGALAGSVTPQSKSAADILEGDKSWGIQTKNAYSSKNKGFTTHSLAVAGKQGF 229
QY
187 KLTSGGINIGLGNK-----FVLHTE--GLYRKSQGYAVPYRYNLRK 225
Db
230 -----GVAIVTHENSITETVHKDALKGVQSYDRFTATTEDQSAFYFMQDECLDGYDKCKT 284
QY
226 LPDSPPRFANGQHRVL-----GWRKRFYRYRTYSRRDOY--GLPA 264
Db
285 SPKRPAFLSTQRETVSVDYTGANRIKPNMKYVESQSM--FLRGYHFSEOHYIGGI-- 339
QY
265 HSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNG 324
Db
340 --FEFTQOKFDI-----RDMTPFAYLRPTDKDLQ-----SRPFYKQDYGAYQHIG 384
QY
325 KPWIDLNKYEYELRAEWKQFPFGFALRVHLNRNDYHDEKAGDAVENFNNQTONARIE 384
Db
385 ---DGRGVKY-----ASGLVFDEHHRKQRVG--TEYIYENKNKAGIID 422
QY
385 LRHOPIGRLKSGWQVQYLGOKGSALSATSEAVKQPMLLDNKVQH----- 428
Db
423 -----KAVLSANQONILDSYMRHTHCSLYPNPKNCRPT 457
QY
429 ---YSPFGVE-----QANW-----DNFT----- 443
Db
458 LDKFYSYTHSDRNRYKEKNHMLQLNLEKIQONWLTHQIAFNGLPDDFTSALQHKDYLTR 517
QY
444 --LEGGVREVKQKASIRYDKALIDRENYKQPLP-----DLGAHQRTARGFA----- 488
Db
518 RVITATASSISEKGEARNGL---QSPVLYPTPKAELVGGDLQNYQKSSNVSDCKVRL 574
QY
489 LSG-NWYFTPOHKLSLTASHQERLPSTQELYAHGKHVATNTEVGNKH-LNKERSNNIEL 546
Db
575 IKGNYYFAARNNAL-----GKTV---DLGLGMRDYVSRTKANESTI 614
QY
547 ALG-YEGDRWQYNLALYRNRFNGNVIYAQTLNDG-RGPKSIEDDSEMKLVRYNQSGADFYG 604
Db
615 SVGKFNFSWNTGIVIKPTWLDLSY--RLSTGFNPSF---ARMYGMRYGKDTDVI 668
QY
605 ABGEIYFKP---TPRYRIGVSDYVRGLKNLPSLPGREDAYGNRPFLAODDQNA PRVP 660
Db
669 GK-----FKPETSNGQFGLKGDGFGNIBSHF-----SNAYRNLIAPABELSKNGTIG 718
QY
661 AARLGPH-----LKASLTDRIDAN-----LDYRVFAONKLA----- 692
Db
719 KGNVGHNAQNAKLVGNVITAOQLDFNGLWKRIPIYGYATFAYNRVKVDQKINAGLASVS 778
QY
693 --RYETRTPGHMLNIGANVRNTRYGEWNWYVKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db
779 SYLFDALQPSRYIIGLYDHPSTNT-----WGIK-----TMTQSKAKSONELLGKRA 825
QY
751 TG--GVNVK 757
Db
826 LGNNSRNK 834

```

Search completed: November 14, 2003, 11:01:15
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:53:22 ; Search time 17 Seconds
(without alignments)
2096.838 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 4036
Sequence: 1 MAQTLKPIVLISILLINPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231	5.7	652	IRGA_VIBCH	P27772 vibrio chol
2	228	5.6	744	HXC1_HAEIN	P44523 haemophilus
3	221	5.5	723	Y262_HAEIN	P44600 haemophilus
4	211	5.2	673	FYUA_YERPE	P46359 yersinia pe
5	210	5.2	673	FYUA_YERPE	P46359 yersinia pe
6	205.5	5.1	676	HMUR_YERPE	O56989 yersinia pe
7	203	5.0	725	HXC2_HAEIN	P45357 haemophilus
8	187.5	4.6	663	CIRA_SCOLI	P17315 escherichia
9	186.5	4.6	687	HEMR_TEREN	P31499 yersinia en
10	184.5	4.6	614	BTUB_SALTY	P37409 salmonella
11	184.5	4.6	810	HPUB_NEIMC	P66949 neisseria m
12	176	4.4	614	YNGD_ECOLI	P06129 escherichia
13	174.5	4.3	700	HPUB_NEIMA	P76115 escherichia
14	169.5	4.2	810	RHTA_RHIME	O91wa2 neisseria m
15	161	4.0	746	FFPA_SCOLI	Q923Q5 rhizobium m
16	160.5	4.0	746	FFPA_SCOLI	P05825 escherichia
17	158.5	3.9	912	TBPI_NEIMB	P44970 haemophilus
18	155.5	3.9	746	FFPA_PSEAE	O50598 pseudomonas
19	155	3.8	908	IFUA_ECOLI	Q06987 neisseria m
20	154	3.8	732	FOXA_ECOLI	P14542 escherichia
21	151.5	3.8	710	FOXA_YEREN	Q01674 yersinia en
22	150.5	3.7	915	TBPI_NEIGO	Q01996 neisseria g
23	144.5	3.6	911	TBPI_NEIME	Q09056 neisseria m
24	141.5	3.5	729	FFUE_ECOLI	P16869 escherichia
25	139	3.4	702	FOXA_SALTY	Q56145 salmonella
26	138.5	3.4	943	LBPA_NEIMB	Q06379 neisseria m
27	134.5	3.3	720	FFPA_PSEAE	P42512 pseudomonas
28	132	3.3	809	PUPB_PSEPU	P38047 pseudomonas
29	131	3.2	635	HTPG_VIBCH	P22359 vibrio chol
30	130	3.2	819	PUPA_PSEPU	P25184 pseudomonas
31	129	3.2	1476	GTFE_STRMU	P08987 streptococ
32	127.5	3.1	944	LBPA_NEIMA	Q91tk4 neisseria m
33	127	3.1	815	FPVA_PSEAE	P48632 pseudomonas

RESULT 1

ID	IRGA_VIBCH	STANDARD;	PRT;	652 AA.
AC	P27772; Q9KUP0;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Iron-regulated outer membrane virulence protein precursor.			
GN	IRGA OR VC0475.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;			
OC	Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1] _			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;			
RX	MEDLINE=93023868; PubMed=1406279;			
RA	Goldberg M.B., Boyko S.A., Butterton J.R., Stoeber J.A.,			
RA	Payne S.M., Calderwood S.B.;			
RT	"Characterization of a Vibrio cholerae virulence factor homologous to			
RT	the family of TonB-dependent proteins.";			
RL	Mol. Microbiol. 6:2407-2418(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=E1 Tor N16961 / Serotype O1;			
RX	MEDLINE=20406833; PubMed=10952301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.R., Read T.D., Tettelin H., Richardson D.,			
RA	Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,			
RA	McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,			
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RT	cholerae.";			
RL	Nature 406:477-483(2000).			
RN	[3]			
RP	SEQUENCE OF 1-151 FROM N.A.			
RC	STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;			
RX	MEDLINE=91072235; PubMed=2174861;			
RA	Goldberg M.B., Boyko S.A., Calderwood S.B.;			
RT	"Transcriptional regulation by iron of a Vibrio cholerae virulence			
RT	gene and homology of the gene to the Escherichia coli fur system.";			
RL	J. Bacteriol. 172:6863-6870(1990)			
CC	- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING			
CC	FERRIC VIBRIOBRACIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS			
CC	V. CHOLERAEE TO EXTRACT IRON FROM THE ENVIRONMENT.			
CC	- SUBCELLULAR LOCATION: Outer membrane.			
CC	- REGULATION OF THE TRANSCRIPTIONAL LEVEL BY IRON			
CC	REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON			
CC	- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.			

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

RX MEDLINE=95350630; PubMed=7542800;
RA Flieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saubek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -I- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC CONCENTRATIONS (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -I- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF
CC H. INFLUENZAE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMSL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U32696; AAC21789.1; -.
DR PIR; B64049; B64049.
DR TIGR; H10113; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_REC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
KW SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 744
FT SEQUENCE 744 AA; 85043 MW; 74D94F72E41AEC31 CRC64;
SQ
Query Match 5.6%; Score 228; DB 1; Length 744;
Best Local Similarity 19.98%; Pred. No. 4e-08;

```


Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 NCBI_TaxID=727;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RC MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Cotton M.D.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RT Science 269:496-512 (1995).
 RN [2]
 RN IDENTIFICATION BY MASS SPECTROMETRY.
 RC MEDLINE=20137488; PubMed=10675023;
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wigf B.,
 RA Gray C., Rountoulakis M.;
 RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
 RT Electrophoresis 21:411-423(2000).
 CC -I- SUBCELLULAR LOCATION: Outer membrane (potential).
 CC -I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; U32712; AAC21927.1; --
 CC PIR; C64058; C64058.
 CC TIGR; H10262;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 DR Outer membrane; Receptor; Signal; TonB box; Complete proteome.
 KW SIGNAL 1 21
 FT SIGNAL 1 21
 FT CHAIN 22 723
 FT SITE 706 723 PROBABLE TONB-DEPENDENT RECEPTOR H10262.
 FT SITE TONB C-TERMINAL BOX.
 SQ SEQUENCE 723 AA; 80775 MW; 91EB3AB0FFEA2984 CRC64;
 Query Match 5.5%; Score 221; DB 1; Length 723;
 Best Local Similarity 19.6%; Pred.No. 1.2e-07;
 Matches 157; Conservative 130; Mismatches 301; Indels 212; Gaps 417;
 QY 11 LSLILLNPLLAQAHETEQVSGLETVTVGKSRPRATSGLLTSTASDKIISGDTLRKA 70
 DB 6 LSLAIITTT-LVTANALAQSVELDSINVIATRPD---SRFAYTPEQSK---DSLSSQA 57
 QY 71 VNLGALGVGFIHASQVGG---ASAPVIRGOTGR-IKVLNHGTEGDMADSPDH-- 124
 DB 58 TSVAADALDIPNDVR--GGSRSTAQPNIRIGSDNRVVDIGVRQNFDA-----HRG 110
 QY 125 ALMVDTALSQQVEITLGRPVTLITVSSGNVAGLVDAVDGKIPEKPNGVSGELGLRLSG- 183
 DB 111 SYFLPMSLIQIEIVIKGPFSSSLMGSGALGVVAAMRTNALDLLKNN---DKFGVKIRGY 167
 QY 184 -NLEKLTSGGINI-GLGNFVILHTEGLYKRSQDYAVPRVNLKRLPDSPPRFANQHRV 241
 DB 168 QTANLSEKDVSVFAANDKFDVLISGPNYADNLRTGKGNKLN--TAYKQFGG--LAK 222
 QY 242 LGWRKRFVRRY--YSDRDQYGLPAHSHEYDDCHADIIWOKSLINKLYOLYPLHLESDV 300

Db 223 FGQIINDANRVELSHRETRFKQTAPSN--:-----EVENELTNEQIT 262

Qy 301 DYDNPGLSCGFHDDDAHAHNGKWDILRNKRYELRAEWKQPPG-:----FEALRVHL 355

Db 263 DQTKK-:-----FHQCKD-:-----DLLPPTQPSERSSEFYKVKTRL 298

Qy 356 NRNDYHHDEKAGD--AVENFF--NNQTONARIELRHQIGLKGSGWQVYLGQKSSALS 410

Db 299 GSVSYLTDOQIPQOSTVFNFYLPDPYLNTHIALYNNKT-:-----IEQRKVSQVK 350

Qy 411 ATSEAVKQPMLLDN--KVQHYSP-FGVEQANNDNFTLEGVVRVEKQ-:-----454

Db 351 DQTKLTRGRINLSSELSHISFVYGVYWR-DKTRTERGTGSDAKFRADPYANSNWT 409

Qy 455 -ASIRYD-:-----KALIDRENYKQPLDPLGAHQRTAKSFALS 490

Db 410 GYVLIHAIPLFGKLVSPSVRYDHYDTGSKTVKYKDNHL-:-----SPATK 454

Qy 491 GNVYFTPOHKLSTASHQE--RLPSTOELVYAGKHVATWTFE-:-----VGNKHLNERS 541

Db 455 LTWIVT--NWLDFTAKYNEAFRAPSQRERFVSGAHFGANTLGLDHNRFVANPNLRPETA 512

Qy 542 NNIEL-:-----ALGVEGRWQYNLALYRNRFNYIYAQTLND-:-----GRGPKS-583

Db 513 KXKEITANLHFDLSLFQGGKFKIEATYFRNDVKDFINLKIFNDAKTSASAGANPNTGAL 572

Qy 584 IEDDSEMKLVRYNSGADFVGAEGEIVFKPTPRYRIGVSGDYVRGELKNLPSLPGREDAY 643

Db 573 LPKNSQVQ-:-----NITNARLSGIELQAOYQ-TERLT-:-----FTNYGSKDKDQS 617

Qy 644 GNRFFAQDDQONAPRVPAPARLGHKLKASLTD-:-----RIDANLDYRVFAQNKLYETRT-P 699

Db 618 G-:-----EALSNTAASRIGVGVNVALYKQKFTVGATWTHYA--AQRVVPKDHSTVP 666

Qy 700 GHMLNLGANYRNTRYGEW 719

Db 667 SYILTDLRAYA--PLKGEW 684

RESULT 4

FYUA_YERPE STANDARD; PRT; 673 AA.

AC P46359;

DT 01-NOV-1995 (Rel. 32, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE Pesticin receptor precursor (IRPC).

GN FYUA OR PSN OR YPO1906 OR Y2404.

OS Yersinia pestis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.

OX NCBI_TaxID=632;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=6769;

RA Rakin A., Heesemann J.;

RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=KIM6;

RX MEDLINE=95204350; PubMed=7896707;

RA Fetherston J.D., Lillard J.W., Perry R.D.;

RT "Analysis of the pesticin receptor from Yersinia pestis: role in

RT iron-deficient growth and possible regulation by its siderophore.;"

RL J. Bacteriol. 177:1824-1833(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=6769;

RA Buchrieser C., Rusniok C., Couve E., Frangeul L., Billault A.,

RA Kunst F., Carniel E., Glaser P.;

RT "DNA sequence of the 102 kbases unstable region of Yersinia pestis.;"

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=CO-92 / Biovar Orientalis;

RX MEDLINE=21470413; PubMed=11586360;

RA Parkhill J., Wren B.W., Thomsen N.R., Titchall R.W., Holden M.T.G.,

RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,

RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Ratrago A.N.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,

RA Felwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,

RA Leatherell S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,

RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Genome sequence of Yersinia pestis, the causative agent of plague.;"

RL Nature 413:523-527(2001).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=KIM5 / Biovar Mediaevalis;

RX MEDLINE=22137863; PubMed=12142430;

RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,

RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,

RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,

RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,

RA Perry R.D.;

RT "Genome sequence of Yersinia pestis KIM.;"

RL J. Bacteriol. 184:4601-4611(2002).

CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE

CC SIDEROPHORE YERSINIAABACTIN.

CC -!- SUBCELLULAR LOCATION: Outer membrane.

CC -!- INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH

CC THE FUR PROTEIN.

CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; Z35104; CAA84487.1; -

DR EMBL; U09530; AAA69906.1; -

DR EMBL; AL031866; CAA21395.1; -

DR EMBL; AJ414150; CAC90722.1; -

DR EMBL; AE013845; AAM85962.1; -

DR PIR; A56148; A56148.

DR PIR; AF0232; AF0232.

DR HSSP; P05825; IREP.

DR InterPro; IPR000531; TonB_boxC.

DR Pfam; PF00593; TonB_dep_Rec; 1.

DR PROSITE; PS00430; TONB DEPENDENT REC_1; FALSE NEG.

DR PROSITE; PS01156; TONB DEPENDENT REC_2; FALSE NEG.

KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;

KW Transport; TonB box; Complete proteome.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 673 PESTICIN RECEPTOR.

FT SITE 30 37 TONB BOX.

FT SITE 657 673 TONB C-TERMINAL BOX.

FT CONFLICT 484 484 G -> D (IN REF. 1).

FT CONFLICT 514 514 R -> G (IN REF. 1).

FT SEQUENCE 673 AA; 73782 MW; 9C39E6010EBCE2C CRC64;

Query Match 5.2%; Score 211; DB 1; Length 673;

Best Local Similarity 21.8%; Pred. No. 5.3e-07;

Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAOTLKPIVLSILLINTPLLAQAHEEQSVGLETVTVGKSRPRATSGLLHTSTASDKI 60

Db 1 MKMTRLYPLALGGLLL--PAIANAQTSQDD---ESTLVVTASKQSSRSA--SANNVSTV 53

Qy 61 ISGDTLFQKAVNLGDALDGV-PGIHASQYGGASAPV-IRGQTGRIRKVLNHHGETGMA 118

Db 54 VSAPELSDAGVTASDKLPVLPLGLNIENSGNMLFSTISLRGVSSAQ-----99

Qy 119 DF-SPDHAINVD-----TALS--QQVEILRGPTVLLYSSGNVAGLVADVADKPIE 165

```

Db 100 DFYNPAVLYVDGVPOLSTNTIQAUTDVSVELLRGPQGLTYGKSAQGGIINIVTQQ-PD 158
QY 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGVIEGVSSRDSYR-SKFNL-----SGPIQDGL-----LYGSVTLRLQVDDGDMINPA 209
QY 220 YRN-----LKRLED-----SPRRFANGQHRAVLGWRKRFYRTYSDR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGMFAASRECTATQDAYVGNW-----258
QY 257 RDQYGLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYNPGLSCGFHDDDD 316
Db 259 -----DIKGRKLSLSDGSPDPYMRCT-----DSQTLSGKYTTDD- 293
QY 317 AHAHANGKFWIDLRNRYELAEWKQ-----PPGFALRVHLNRNDYHDEK-----365
Db 294 -----WV-----FNLSAWQOQHYSTFPB-GSLIYNMPQR-WNQDVQELRAAT 335
QY 366 AGDA--VENFFNQTQNAIRIELRHQPIGRKLSGWSGVYLGQKSSALSATSSEAVKQPMLLD 423
Db 336 LGDARTVDMVFGLYRQNTN-----EKLNSAYDMETMPLYLSSTGYTTAETLAA-----382
QY 424 NKVQHSFFGVEQANW---DNFTLEGVVRVEKQASIRYDKALIDRENKYKPLDLGAH 480
Db 383 -----YS-----DLTWHLTDREDFDGGVRFSDKSSSTQYHGSMLG-----NPFQDQGS 426
QY 481 RQTARFALSNGWYFTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLNKR 540
Db 427 NDDQVLQGLSAGVMLTDMRW-YTRVAQYKFSGNYI-----VPTAGLDARFFVAEK 477
QY 541 SNNIELALYEGDRQWYNALYRNFPGYIYAQTINDGRGPKSTEDDSEMLKVRYNOSGA 600
Db 478 SINEYLGTRYE---TADVTLOAATF---YTHTKDQMLYSGPVVRMQTLS-----NAGKA 524
QY 601 DFYGAAGEIYKPTPRYRIGVSGDVVRGRKLNPLSGREDAYGNRPFIADQDNAPRVP 660
Db 525 DATGVELEAKWFAFGSWDINGVIRSEFTN-----DSELYHGNR-----VFPVP 570
QY 661 AARLGHFKASITDRIDANLQYRVPAQNKALARYETRTFGHHMLNGLANYRNTRYGB-- 718
Db 571 RYAGSSVNGVIDTRYGA-----LMPFLAVNLVGHYFD-GDNQLRQGTATLD 618
QY 719 ---W-----NWKYKADNLLNQ---SVYAHSSFLSDTPCMGRSFTGGVNVK 757
Db 619 SSLGQWATERMNLISVVDNLFDREYETGYMGNGSSAVAQVNMGRVGINTR 669

RESULT 5
FYUA YEREN STANDARD; PRT; 673 AA.
ID FYUA YEREN
AC P46360;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticin receptor precursor (IRPC) (IPR65).
GN FYUA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51871 / WA-314 / Serotype O:8;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RL factor with dual function."
RL Mol. Microbiol. 13:253-263(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1223-75-1, 8081 / Serotype O:8, YE737, and 1209-79;
RA Rakin A., Heesemann J.;

```

```

RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC SIDEROPHORE YERSINIACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z29675; CAA82773.1; -
CC EMBL; Z35485; CAA84620.1; -
CC EMBL; Z35486; CAA84621.1; -
CC EMBL; Z35487; CAA84622.1; -
CC EMBL; Z35496; CAA84630.1; -
CC PIR; S60142; S60142.
CC HSP; P05825; 1PEP.
CC InterPro: IPR00531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rec; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
CC Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
CC Transport; TonB_box.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 673 PESTICIN RECEPTOR.
CC SITE 30 37 TONB_BOX.
CC SITE 657 673 TONB_C-TERMINAL_BOX.
CC FT VARIANT 362 362 N -> D (IN STRAINS 8081, YE737 AND 1209-
CC 79).
CC SQ SEQUENCE 673 AA; 73682 MW; 1C20E0352FAD4DCC CRC64;

Query Match 5.2%; Score 210; DB 1; Length 673;
Best Local Similarity 21.7%; Pred. No. 6.2e-07;
Matches 180; Conservative 95; Mismatches 320; Indels 236; Gaps 43;

QY 1 MAQTLLKPIVLISILLINTPLLAQAHETESQVGLTIVTVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTRLYPLALGGLLI--PAIANAQTQQDESTLEVT-ASKQSSRSAS---ANNVSSTV 53
QY 61 ISGDTLRQKAVNLGALDGV-PGIHASQVGGGASAPV-IRGOTGRRIKVINHHGTGDMA 118
Db 54 VSAPELSDAGVTASDKLPVLPGLNINENGNMFTSILRGVSSAQ-----99
QY 119 DF-SPDHAIWVD-----TALS--QQVEILRGPVTLTYSSGNVAGLVADVADKTIPE 165
Db 100 DFYNPAVLYVDGVPOLSTNTIQAUTDVSVELLRGPQGLTYGKSAQGGIINIVTQQ-PD 158
QY 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGVIEGVSSRDSYR-SKFNL-----SGPIQDGL-----LYGSVTLRLQVDDGDMINPA 209
QY 220 YRN-----LKRLED-----SPRRFANGQHRAVLGWRKRFYRTYSDR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGMFAASRECTATQDAYVGNW-----258
QY 257 RDQYGLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYNPGLSCGFHDDDD 316
Db 259 -----DIKGRKLSLSDGSPDPYMRCT-----DSQTLSGKYTTDD- 293
QY 317 AHAHANGKFWIDLRNRYELAEWKQ-----PPGFALRVHLNRNDYHDEK-----365
Db 294 -----WV-----FNLSAWQOQHYSTFPB-GSLIYNMPQR-WNQDVQELRAAT 335
QY 366 AGDA--VENFFNQTQNAIRIELRHQPIGRKLSGWSGVYLGQKSSALSATSSEAVKQPMLLD 423
Db 336 LGDARTVDMVFGLYRQNTN-----EKLNSAYDMETMPLYLSSTGYTTAETLAA-----382
QY 424 NKVQHSFFGVEQANW---DNFTLEGVVRVEKQASIRYDKALIDRENKYKPLDLGAH 480

```

```

Db 383 -----YS-----DLTWLTDREDDIGGVRFSDKSTQYHGSMLG-----NPFQDQCKS 426
Qy 481 RQTARFALSGNWYFTPQHKLSLTASHOERLPSTQELVAHGKHVATNTFEVGNKHLNKR 540
Db 427 NDDQVUGLSAGYMLTDHVRV-YTRIAQYKESGYNI-----VPTAGLDARFFVAEK 477
Qy 541 SNNIELALGVGDRWQYNALVRNPGNVIYAQTLNDGRGPKSIEDSEMGLVRYNQSGA 600
Db 478 SINVELGTRYE-----TADVTLOAATF--YTHTKDMQLYSGPVGMQTLT-----NAGKA 524
Qy 601 DFYGAEGEIVFKPTPRYRIGVSDVVRGRKKNLPSLPGREDAYGNRPFIADQDQNAVRVP 660
Db 525 DATGVLEAKWRPAPGSHWDINGNVRSEFTN-----DSLEYHGNR-----VPFVP 570
Qy 661 AARLGFHLKASLDRDANLDYVRFAONKRLARYETRTPGHHMLNLGANYRNRTRYGE-- 718
Db 571 RYAGAGSSVNGVIDTRYGA-----LMPRLAVNLVGPYHFD-GDNQLRQGTATYATLD 618
Qy 719 ----W-----NWYVKADNLNLO--SVYAHSSFLSDTPQMGSRFTGGVNVK 757
Db 619 SSLGWQATERINISVHVDNLFDERYTYGYMNGSSAVAQVNNGRVGINTR 669

RESULT 6
HMUR YERPE
ID HMUR YERPE STANDARD; PRT; 676 AA.
AC Q56989;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemin receptor precursor.
GN HMUR OR YPO0283 OR Y0543.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM6;
RX MEDLINE=96386041; PubMed=9026634;
RA Horning J.M., Jones H.A., Perry R.D.;
RT "The hmu locus of Yersinia pestis is essential for utilization of
RL free haemin and haem-protein complexes as iron sources.";
RN Mel. Microbiol. 20:725-739(1996).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RN Nature 413:523-527(2001).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=42137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Ferherton J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RN J. Bacteriol. 194:4601-4611(2002).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATING SIDEROPHORE THAT ALLOWS
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U60647; AAC64866.1; --
DR EMBL; AJ414141; CAC89146.1; --
DR EMBL; AE013655; AAM84131.1; --
DR PIR; AG0035; AG0035;
DR PIR; T12069; T12069.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 676 HEMIN RECEPTOR.
FT SITE 44 51 TONB BOX.
FT SITE 659 676 TONB C-TERMINAL BOX.
SQ SEQUENCE 676 AA; 74230 MW; 84ED731CB914ACD3 CRC64;

Query Match 5.1%; Score 205.5; DB 1; Length 676;
Best Local Similarity 19.3%; Pred. No. 1.3e-06;
Matches 164; Conservative 110; Mismatches 282; Indels 293; Gaps 37;

Qy 11 LSTLLINT-PLLAQAHEEQSVGLTETVVGVKSRPRATSGLLHTSTASDK----- 59
Db 14 LSLAIACTLFLATQA-----ADTTTQTSSKXSTDTWVVTATGNERSSEAFPMVMT 65
Qy 60 IISGDT-LQKAVNLGDALDVGPIHAGSYGGAGAFVIRGOTGRRIKVLNHHGTGDM- 117
Db 66 VIEGNAFTSQTATAADMLRQVPLTVTG-----SGRTNQDVVMRGYGHQGVLT 115
Qy 118 -----ADFSPDHAIMVDTALSOQVEILRPVTLTYSSGNVAGL-----VDVADGKIP 164
Db 116 LVDGVRGQDTGHLNSTFLDPAIVKRIEIVRGPAALYGGALGGVIAVETVDAAD---- 171
Qy 165 EKPENGVSCEGLRLSSGNLEKLTSGGINIGKGNFVLHTE-----GLYRK 211
Db 172 --MLQPGQNS--GYRVIS-----SAATGDSHSGFGLGASAFGTDDGLGILSFGTRDIGNRQ 223
Qy 212 SGDYAVPR-----YRNLEKRLDPSRRFANGQHRVGLW 244
Db 224 SNGFNAPNDETI SNVLAKGTWQIDSIQSLSANLRYNNSAIEPNQPSAPSSTN-VMTN 282
Qy 245 RKRFRYRTYSDRDQYGLPAHSHEYDDCHADIIWOKSLINKRYLQLYPHLLTEEDVDYDN 304
Db 283 RSTIQ-----DAQRYNIRKPLDQEWLNATAQVYSEVEINAR-----PGSABEGREQT 333
Qy 305 PGLSCGFHDDDDAHAAHNGKPMIDLRNKRYELRAEWKOPFPFPAALRVHL---NRNDYH 361
Db 334 EGVK-----LENRTLF-----IESPASHLLTYGTETK 362
Qy 362 HDEKAGDAVENF-----FNNQTONARIELRHQPIGRKSGWGVQLGOKSSALSATSEA 415
Db 363 QEOTPGGATESFPQAKIRFSGWLQDEITRLDPLVSILAGT----- 403
Qy 416 VKQFMLLDNKVQHSYFPGVEQANWDNFTLEGGRVVEKOKASIRYDKALIDRENYTKQPLP 475
Db 404 -----RYDYSGSSDGY-----A 416
Qy 476 DLGNHRCRTARSPALSGNWYFTPQHKLSLTASHQE--RLPSTQELVAHGKHVA----- 525
Db 417 DVDADKWSRG-AIS-----ITPTDMLMFGSAQAFAFTMGEMYNDSKHFAPIRPGLT 471
Qy 526 -TNTFEVGNKHLNKRNNIELALGYE-----GDRWOYNLALVRNPGNVIYAQTLND 577
Db 472 LTN-YWVPENLKPETNETQCYGFGRLPFDLLMAEDDLQFKVSYFDTRAKDYISTREV--- 527

```

```

QY 578 GRGPKSIEDSEMMLVRYNSGADFYGAEGEYFKTPRYRIGVSGDYVGRGLKNLPSLP 637
Db 528 -----DQAMTTSVNIQAKINGDASMSYK-TALFNWDLAYNRTRGKNQ----- 573
QY 638 GREDAYGNRPFFIAODQONAPR-----VPAARLGFHLKASLTDRIIDANLDYRVFAQNL 691
Db 574 --TDEW-----LDTINPDTVTSIVDPVANGSF-----SVGMIGTEFA-NRS 611
QY 692 ARYETETP--GHEMLNLGANYRNTRYGWNWVVKADNLNLOSVAHSSFLSDTPQMGRS 749
Db 612 SRVSSSTPQAGYGVNDFFYSYKQOAFKGMTTMLLGNVFKEYYTPQGI-----PDGR- 666
QY 750 FTGGVNVKF 758
Db 667 -----NVKF 670

RESULT 7
HXC2_HAEIN STANDARD; PRT; 725 AA.
AC P45357;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Heme/hemopexin utilization protein C precursor.
GN HXUC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC STRAIN=D42 / Serotype B;
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yergev R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
RL heme/hemopexin by Haemophilus influenzae type b.";
RT J. Bacteriol. 177:2644-2653 (1995).
CC -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC CONCENTRATIONS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
CC H. INFLUENZAE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09840; AAA87059.1; -.
DR PIR; A57148; A57148.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB Box; Signal; Receptor.
FT SIGNAL 1 19 OR 21 (POTENTIAL).
FT CHAIN 20 725 HEME/HEMOPEXIN UTILIZATION PROTEIN C.
SQ SEQUENCE 725 AA; 80838 MW; FC7886E020CB5BCE CRC64;

Query Match 5.0%; Score 203; DB 1; Length 725;
Best Local Similarity 19.8%; Pred. No. 2.1e-06;
Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;

QY 11 LSLILNTPLLAQAHETEQSVGLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLRQKA 70
Db 6 LSLAIATT--LVTNALAQSVELDSINVIATRDP---SRFATPEKQSK---DSLILSKQA 57

```

```

QY 71 VNLGDALDGVPGIHASQYGGG---ASAPVIRGOTGR--IKVLNHHGETGMADFPD-- 124
Db 58 TSAAALEDIPNVDIR--GGSRSIAQPNIRGSDNRVQVQIDVRQNFLLA-----HRG 110
QY 125 AIVVDTALSQQVILRGPVTLTYSSNGVAGLVADVADGKIPEKMPENGVSGBELGRLSSG- 183
Db 111 SYFLPMSLLQIEIVIKGPSSSLWGSALGGVAVMRTENALDLLKN--DKFGVKIRQGY 167
QY 184 -NLEKLTSGGINI--GLGKNFVLHTEGLYKSGDYAVPRYRNLRKLDPSPRRFANGQHRV 241
Db 168 QTANNLSESDVSFAANDKFDVLISGFYNADNLRTGKGNKLN--TAYKQFGG--LAK 222
QY 242 LGWRKFFYRT--YSRRDQYGLPAHSHYEDDCHADIIWQKSLINKRYLQLYPHLLTBEV 300
Db 223 FGWQINDANRVELSHRETRFKQTAPS-----NEVENLTNEQITDOIR 266
QY 301 DYDNPGLSGCFHDDDDAHAAHAKGPMWIDLRNKRYELRAEWKQPPGFALRVHLNRNDY 360
Db 267 EFHKP-----NNGSP-----PKAKPSQEEF-----YSGVKTFRGVS 299
QY 361 HHDEKAGD--AVENFF--NNQTONARIELRHQPIGRKSGWGVQYLGQKSSALSATSEA 415
Db 300 LTDQIIPDQSTVFNYLTPDNVYLNTHIALYNNKT-----IEKQKRVSGVKDQTKL 351
QY 416 VKQPMLLDN--KVQHSYF--EGVEQANWDNFTLEGV-----RVEKQKA----- 455
Db 352 TTRGINLRNSELSHISFVYGVGYMR--DKIRTERGTNNKDAQFRADPNANSNTTGVYLI 410
QY 456 -----SIRYD-----KALIDRENYKQPLDGLGAHROTARSAFALSGNMYF 495
Db 411 AHIFLFGKLLLSPSVYDHTVTSKTVYKDNHL-----SPATKLTWIV 455
QY 496 TPQHKLSLTASHOB--RLPSTQELYAHGKHVAT-----NTFEVGNKHLNKRSNIE 545
Db 456 T--NWLDFATKYNEAFRAPSQMERFVSGSHFGTSLGRNEINKF--VANPNLRPETAKKE 512
QY 546 L-----ALGYEGDRWQYNLALYRNFVNYIAQTLDNGRGRKSIEDSEMMLVRYNQS 598
Db 513 ITANLHFDLSLFKQGDKFKEATYFRNDVKDPINLKFINDAKTINTNASASA-----GA 564
QY 599 GADFVGAEGEYFKPT--PRYR-----IGVSGDYVRGRL---KNLPSLPGRDAYCN 645
Db 565 GA---GANPGALLPTKSOYQONITNARLSGLIELAQYQTERLTFTNYGSGTKGDKDSG- 620
QY 646 RPIAQQDDQNAAPVPAARLGFHLKASLTD--RIDANLDYRVFAQNLKARYETET-POH 701
Db 621 -----EALSNTAAASKIGVGVNYALVKDKFTVCATVTHYA--AQRVVKDHSVTPSY 670
QY 702 HMLNLGANYRNTRYGEW 719
Db 671 ILTDLRATYA--PLKGEW 686

RESULT 8
CIRA_ECOLI STANDARD; PRT; 663 AA.
AC P17315;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Colicin I receptor precursor.
GN CIRA OR CIR OR FEUA OR B2155.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=89123100; PubMed=2644220;
RA Nau C.D., Konisky J.;
RT "Evolutionary relationship between the TonB-dependent outer membrane
RT transport proteins: nucleotide and amino acid sequences of the
RT Escherichia coli colicin I receptor gene.";

```



```

RESULT 9
ID HEMR YEREN STANDARD; PRT; 687 AA.
AC P31459;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hemin receptor precursor.
GN HEMR.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RX MEDLINE=93049186; PubMed=1425573;
RA Stojiljkovic I., Hantke K.;
RT "Hemin uptake system of Yersinia enterocolitica: similarities with
other TonB-dependent systems in Gram-negative bacteria.";
RL EMBO J. 11:4359-4367(1992).
RN [2]
RP REVISIONS.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA Stojiljkovic I.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -!-
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -!-
DR EMBL; X68147; CAA48250.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rc; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Outer membrane; Iron transport; Transpor; TonB box; Signal;
Receptor.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 687 HEMIN RECEPTOR.
FT SITE 44 51 TONB BOX.
FT SITE 670 687 TONB C-TERMINAL BOX.
SQ SEQUENCE 687 AA; 75226 MW; 69092EP97F5DC08A CRC64;

Query Match 4.6%; Score 186.5; DB 1; Length 687;
Best Local Similarity 19.2%; Pred. No. 2,8e-05;
Matches 164; Conservative 94; Mismatches 262; Indels 333; Gaps 39;

QY 8 PIVLSILLINTPELLAQAHTEGVSGLVTVTVVKGRPRATSGLLHTSTASDK-----59
Db 13 PLSLAI-ACTL5LAVQAADTS-----STQT--NSKKRIADTWVVTATGNERSFEAPMM 63
QY 60 --IISGDT-LRQKAVNLGDALGVPCIHASQYGGGASAPVIGOTGRRIKVLNHHGETGD 116
Dd 64 VTVWEADTPTSETATSDMLRNIFGLTVT--GSGR-----VNGQD-----VTLRGYKQGV 113
QY 117 M-----ADFSFPHAIMVDTALSSQVEILRGPVTVLLYSNGVAGL-----VDVADGK 162
Dd 114 LTLVDIGRGTTGHLNSFLDPALVKRVEIVRGSALLYSGALGGVISYTVDAADLL 173
QY 163 IPEKMPENGVSGLGLRUSGSMLEKLTSGGINIGLGNFVLHTE-----GLY 209
Dd 174 LPGQ-----NSGRVYVSA-----AATGDHSGFLGASAFGRDTDVDDGILSPGTRDIGNI 221

```

```

QY 210 RKSGDYAVP-----RYRNKLKL-PDSPPRFANGQRAVL 242
Dd 222 RQSDGFNAFNDETTISNVLAKGTWRIDQIQSLSANLRYNNNSALEPKNPQTSAASTNLMT 281
QY 243 GWRKEFYRTYSDDRDQYGLPAHSHEYDDCHADIIWOKSLINKRYLQLYPHLLTEDVDY 302
Dd 282 D-RSTIQ-----DAQLKNIKPLDQEWLNATRAQVYSEVINAR-----PQTPPEGRKQ 331
QY 303 DNPGLSCGFHDDDDAHAHNGKPMWIDLNRKRYELRAEMWKQPPGFEALRVHL---NRND 359
Dd 332 TTKG-----GK--LENKTRLFT-----DSPSHLLTYGTEA 360
QY 360 YHDEKAGDAVENF-----FNNQTONARIELRHOPIGELKSGWGVQYLGOKSSALSATS 413
Dd 361 YKQEQTPSGATESFPQADIRFGSGWLQDEITLRLDLPVSILAGT-----403
QY 414 EAVKQPMLLDNKVQHYVFFGVEQANWNTLLEGGVVRVEKQASIRYDKALIDRENYKQP 473
Dd 404 -----RYDNYRGSSEGY-----415
QY 474 LPDLGAHRQTARSFALSGNWYFTPOHKLISLTASHOE--RLPSTQELYAHGKHVATN---527
Dd 416 -ADVADAKWSSR-----GAVSVTPTDMLMLFGSYAQAFRAPTMGEMYNDSKHFSSMIMGN 469
QY 528 ---TFEYGVNKHINKERSNNIELALGYEGDRWQYNLALYRNRFGNYIYAQTINDRGPKSI 584
Dd 470 TLTNYWVFNPLKPKETNETQYGFGL-----RFNDLMMMA-----503
QY 585 EDDSMKLVRYNOSGADFYGABGEIYFKPTPRYRIGVSGDYVVRGLKNLPSLPGRDAYG 644
Dd 504 EDDLQFKASYFDTNKDYIST-----GVTDGFGF-----PG-----535
QY 645 NRPFTAQDDQNAAPRVPAAALGFHLKASLTDRIDANLDYRVFPAQNKLAARYETRTPGHML 704
Dd 536 -----GLYCKNCST--YSTNIDRAKINGWDATMTYQ-----WF 568
QY 705 NLGANYRR-----NTRYGEMWVYKADNLLN--QSVYAHSSFL-----SD 742
Dd 569 NLGLAYNTRGKNQNTNEMLDITNPDTVTSTLDVPVANSFGAVGWIGTFADRSSRVSSG 628
QY 743 TPQMGSRSTGGVN 755
Dd 629 TPQAGY-----GVN 637

RESULT 10
ID BTUB SALTY STANDARD; PRT; 614 AA.
AC P37409;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vitamin B12 receptor precursor.
GN BTUB OR STM4130.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=93079446; PubMed=1448622;
RA Wei B.Y., Bradbeer C., Kadner R.J.;
RT "Conserved structural and regulatory regions in the Salmonella
typhimurium btub gene for the outer membrane vitamin B12 transport
protein.";
RL Res. Microbiol. 143:459-466(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

```


Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
 "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
 Nature 413:852-856(2001).
 CC -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M89481; AAA27031.1; -;
 DR EMBL; AB008893; AAL22968.1; -;
 DR StyGene; SGI0030; btub.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Cobalt transport; Transport; TonB box; Signal;
 KW Receptor; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 614
 FT SITE 26 33
 FT SITE 597 614
 FT CONFLICT 81 81
 FT CONFLICT 134 134
 FT CONFLICT 183 183
 FT CONFLICT 280 280
 FT SEQUENCE 614 AA; 68525 MW; 9F51F601A615FD62 CRC64;
 SQ
 Query Match 4.6%; Score 184.5; DB 1; Length 614;
 Best Local Similarity 21.5%; Pred. No. 3.2e-05;
 Matches 140; Conservative 83; Mismatches 235; Indels 193; Gaps 35;
 29 QSVGLTETVVGKSRPRATSGLLHTSTASDKIISGDTLFRQKAVNLGDALDVGPGIHASQY 88
 21 QDTPDLVVTANRQQPSASVAPVTI---VTRQDIERQWSTVNDVRLRPGVDIAGS 77
 89 GGGA--SAPVIRGOTGRRIKV-----LNHGETGMDADFPDHAIMVDLTALSQQVEILR 140
 78 GGAGQNSSIFIRGTNSSHVLVLIDGVRNLAVGSGS-ADLS-----QFPVSLVQRIEYIR 131
 141 GPVTLTLYSGNVAGLVADVADGKIPEKMPENGVSGLGLRSLSGNLEKLTSG-GIN----- 194
 132 GPRSAIYGSDAIGGVNI-----ITRDNFGT--ELTAGWGSNSYQNY 172
 195 -----IGLKNFVLHTEGLYRKGDYAVPYRNLKRLPDSFRFPANGQRAVLGWRKR 247
 173 DISTQQQLGENTRATLIGDYETKGFVDVAVKGTGMQAPDRDGFSLKTYGAL----- 226
 248 FVRRYSRRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQIYPHLLTEEDVDVNDPLG 307
 227 --EHTFSRWSGF---VRGSGYDN-----RTDYD----- 250
 308 SCGFHDDDDAHAAHNGKFWIDLRNKRVELRAEWKQPPFPFALRVHL-----NRNDYHH 362
 251 -----AYSPGSLIDTR-KLYSQSDWAGLHFG-ERISQLVSSVSHSKDNYN 297
 363 DKAG-----DAVENFNNQTONARIELRHQPIGRKSGWGVYLGQKSSALTSATSE 414
 298 DPHYGRYDTSATLDEMKG-YNVQWTNS-VVVGHGNVGA-----GVDW--QKQTTTPEGTY 348
 415 AVKQPMLDNKKVHYSFVEQVQANWNTFTLEGVRVVEKQKASIRYDKALIDRENTYKQPL 474
 349 V---PEGYDQRTNGTVYLTGLQILG--DFTLEAARSDDNS----- 383

QY 475 PDLGAHROTARSFALSGNMYFTPOHKLSTASHQRLPSTQBYLAHGKHVATNWFVGNK 534
 DB 384 -QFGRH---GTWQTSAGMEFIEGYRFTASYGTSYKAPNLGLQYGY-----YGNP 428
 QY 535 HLNKRSNNIELALGYE---GDRQYNLALYRNFNGVYIYAQTINDGRGPKSIEDSEM 590
 DB 429 NLNPKSKOWEGA--FEGLTAGVSWR--ISGYRN-----DIND-----MIDYDHL 470
 QY 591 KLVRYNQSGADFYGAEGEYFKPTPRYRIGVSGDYV--RGRLNKLPSPLPGR 639
 DB 471 QKY-INEGKARIKIGIEATNFDTGTLTHT-VSYDYVDARNAITDTP-LPRR 518
 RESULT 11
 HPUB NEIMC
 ID HPUB NEIMC STANDARD; PRT; 810 AA.
 AC P96949;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin-haptoglobin utilization protein B precursor.
 GN HPUB.
 OS Neisseria meningitidis (serogroup C).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=135720;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DNM2 / Serogroup C / Serotype 2a;
 RX MEDLINE=97206152; PubMed=9157245;
 RA Lewis J.A., Gray E., Wang Y.-P., Roe B.A., Dyer D.W.;
 RT "Molecular characterization of hpuAB, the haemoglobin-haptoglobin-
 utilization operon of Neisseria meningitidis.";
 RL Mol. Microbiol. 23:737-749(1997).
 CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
 CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (potential).
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U73112; AAC44893.2; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Transport; TonB box; Signal; Receptor.
 FT SIGNAL 1 22
 FT CHAIN 23 810
 FT HEMOGLOBIN-HAPTOGLOBIN UTILIZATION
 FT PROTEIN B.
 FT SITE 793 810
 FT TONB C-TERMINAL BOX.
 SQ SEQUENCE 810 AA; 90626 MW; 0D08DD79DA9CB817 CRC64;
 Query Match 4.6%; Score 184.5; DB 1; Length 810;
 Best Local Similarity 20.9%; Pred. No. 4.8e-05;
 Matches 190; Conservative 114; Mismatches 346; Indels 257; Gaps 49;
 7 KPVLVILLINT-PULLAQAHETEQSVGLTETVVGKSRPRATSGLLHTSTASDKIISGPT 65
 6 KPVLAAVAIAQAFPAADPAQSAQTLEITVTGTHK-----TQKLGEEKIRKT 56
 66 LRQKAVNLGDALDGV---PGIHASQY-GGASAPVIRGOTGRRIKV----- 107
 57 LDKLLVN--DEHDLVRYDPGIGSVVEGGRAGSGFTIRGVDKDVAINVDGLAQAESRSSE 114
 108 -----LNHHGETGMDADFPDHAIMVDLTALSQQVEILRGPVTLTLYSGNVAGLV 156

```
Db 115 AFQELFGAYGNFNAWNTSEPNFS-----EVVITKGADSLKSGSGALGAV 161
Qy 157 DV----ADGKIPKWPEN-GVSGELGLRLSSGNLEKLTSGGINICL-----GK-- 199
Db 162 NYQTKSASDVSEDKPYHLGKGG-SVGKNSQKSFSSITAAGRLFGDLALLVYTRFRGKET 220
Qy 200 -----NFVLHTEGLRKSGDYAVP-RY-----RNLRKLPDPSRRFA----- 234
Db 221 KNRSTEGNEIKNDGYVNPYNTDTGGPKSYLTYPATGVARSQDPQEWVNKSTFLKLYNF 280
Qy 235 NGQHRVGLWKRKFRYRTYSRRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHL 294
Db 281 NDQNR--IGW--IFESRTR-----FTNLSN-----LWTGTTSAATGD-YRH- 320
Qy 295 LTBEDVDP-DNPGUSCGFHDDDDAHAAHNGKWPIDLNRKRYELRAE---WKQFPFGFEA 350
Db 321 --RQDVSRRSSGVE--YKNELEHG-----PWDSLKRLYDKQRIDMNTWTWIDPKNYD 369
Qy 351 LRVLNRNDYHDEKAGDAVENFFNQTON-----ARIELRHQPIGRKSGNGVOY---L 402
Db 370 LR-GINSEVYHS-----FHRIRQNTAQTADFE---KQLDFSKAVMAAQYGLGG 414
Qy 403 GQKSSALS-----ATSEAVKQPMLLDNKVQHYSGFVGEQANWDFTEGGVR 449
Db 415 GRGNANSYDYFVKLYDPKILTSNOAKITMLIENR-SKYKF-----AYWNVHFLGND 468
Qy 450 VEKQASIRYDK---ALIDRENY---YKQPLPDLGAHQRTAR-SFALSGNMYFTPOHKLS 502
Db 469 RFRLNAGIRYDKNSSAKDDPKYTTAIRGQIPHLGSERAHAGFSVGTGDFWRTKHLHL 528
Qy 503 LTSHQERLPSTQELYAHGKHVATNTEVGNKHNKERSNNIELALGVEGRWYNALY 562
Db 529 AKYSTGRAPTSDETWLLFPH--PDFYKAPNLKAEKARNWEGLAGSKAGNPKLSGF 586
Qy 563 RNRENGYI-----YAQTLNDGRGPKSTEDSEMKLRYNOS-----G 599
Db 587 KTKYRDFIELTYMGVSSDNPNKPYA-PLSDGTALVS-----SPVWQNRSAWVK 638
Qy 600 ADFVGAGEIYFKPTPRYIGVSGDYVRGRUKNLPSPGREDAYGNRPFFIAQDDQNA-- 657
Db 639 LEFNGTNWLSIGLPGQTHAGVNVSYIKGKA---QTNGQE-----TPINALSPWSAVYN 690
Qy 658 ---RVPARGLFHLKASIT-----DRIDANLDYRVYPAQNKLARYETRTPGHEMLNG- 707
Db 691 LGYDAPSKRWGINAYATAAKESDTHSNDLNNPWPYAKHASKAYTLFDLSAYLNIKG 750
Qy 708 -----ANYR-RNTRYGEW-----NMVVKADNLNQSVYAHSSFLSDTPQWGRSFT 751
Db 751 QVTLRAAAYNITNKQYVTWESLRSIRBEFTVNRVDNKTTHAGIQRTS-----PGRSYN 803
Qy 752 GGVNVKF 758
Db 804 FTIBAKF 810
```

RESULT 12

```
BTUB_ECOLI
ID BTUB_ECOLI STANDARD; PRT; 614 AA.
AC P06129;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vitamin B12 receptor precursor.
GN BTUB OR BFE OR CER OR DCRC OR B3966.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130824; PubMed=3882670;
RA Heller K., Kadner R.J.;
```

```
RT "Nucleotide sequence of the gene for the vitamin B12 receptor protein
RL in the outer membrane of Escherichia coli.";
RN J. Bacteriol. 161:904-908(1985).
RP [2]
RC SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RL region from 89.2 to 92.8 minutes.";
RN Nucleic Acids Res. 21:5408-5417(1993).
RC [3]
RX SEQUENCE OF 1-5 FROM N.A.
RA MEDLINE=91154132; PubMed=1999392;
RA Gustafsson C., Lindstroem P.H., Hagervall T.G., Esberg B.,
RA Bjork G.R.;
RT "The trmA promoter has regulatory features and sequence elements in
RL common with the rRNA P1 promoter family of Escherichia coli.";
RN J. Bacteriol. 173:1757-1764(1991).
RC [4]
RX SEQUENCE OF 456-614 FROM N.A.
RA STRAIN=RDD020;
RX MEDLINE=93106943; PubMed=8093236;
RA Dougherty T.J., Thanassi J.A., Pucci M.J.;
RT "The Escherichia coli mutant requiring D-glutamic acid is the result
RL of mutations in two distinct genetic loci.";
RN J. Bacteriol. 175:111-116(1993).
RC [5]
RX SEQUENCE OF 21-32.
RA STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RL in the genome of Escherichia coli K-12.";
RN Electrophoresis 18:1259-1313(1997).
RC [6]
RX MUTAGENESIS OF TONB BOX.
RA MEDLINE=90078094; PubMed=2687240;
RA Gudmundsdottir A., Bell P.E., Lundrigan M.D., Bradbeer C.,
RA Kadner R.J.;
RT "Point mutations in a conserved region (TonB box) of Escherichia coli
RL outer membrane protein Btub affect vitamin B12 transport.";
RN J. Bacteriol. 171:6526-6533(1989).
CC -1- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10112; AAA23524.1; --
CC EMBL; U000006; AAC43072.1; --
CC EMBL; AE000471; AAC76948.1; --
CC EMBL; M57668; -- NOT ANNOTATED_CDS.
CC EMBL; L14556; AAA23676.1; --
CC PIR; A65204; QRECBT.
CC EcoGene; EG10126; btuB.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rc; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Cobalt transport; Transport; TonB box; Signal;
CC Phage recognition; Receptor; Complete proteome.
CC SIGNAL 1 20
CC CHAIN 21 614 VITAMIN B12 RECEPTOR.
CC SITE 26 33 TONB BOX.
```

```

FT SITE 597 614 TONB C-TERMINAL BOX.
FT MUTAGEN 28 30 L->P: INACTIVATE UPTAKE.
FT MUTAGEN 30 30 V->G: INACTIVATE UPTAKE.
FT CONFLICT 162 162 A -> G (IN REF. 1).
FT CONFLICT 377 377 A -> R (IN REF. 1).
SQ SEQUENCE 614 AA; 68407 MW; AB43CC46A991FF95 CRC64;

Query Match
Best Local Similarity 19.8%; Pred. No. 0.00013;
Matches 132; Conservative 78; Mismatches 211; Indels 246; Gaps 31;

Qy 29 QSVGLTETVTVGK--SRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGALDGVFGIIAS 86
Db 21 QDTSPTLVVTVANRFEQPRST--VLAPTIV--VTRQDIDRWQSTSVNDVLRRLRGVDIT 75

Qy 87 QVGGGA--SAPVIRGOTGRRIKV-----LNHHGETGDMADFSPDHAIMVDFTALSOQVEI 138
Db 76 QNGGSGQLSSIFIRGTNASHVLVDIGVRLNLAGVSGS--ADLS-----QPPALVQRVEY 129

Qy 139 LRGPVTLTYSNGVAGLVADVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGINIGLG 198
Db 130 IRGPRSAVYGSDAIGGVNIIITR-----DEPQTEISA-----GWG 165

Qy 199 KNFVLHTEGLYKSGDYAVPRYRL-----KRLPDSPPRFANGQRAVLGWRKRYRRTY 253
Db 166 SN-----SYQNDVSTQQQGDKTRVTLG DY----- 192

Qy 254 SDRRDQVGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDVNDPGLSCGF-- 311
Db 193 -----ATHGYD-----VVAYGNTGTQAGTDND-----GFLS 219

Qy 312 -----HDDDA-----HAHANGKEPIDLR--NKRKYELRAEWKOP 344
Db 220 KTLYGALHNFDTAWSGFVRGYDNRITDYAYSPGSLDTRKLYSQSDAGLRYNGE 279

Qy 345 PGFPEALRVHLNRNDVHHDEKAG-----DAVENFFNNTQONARIELRHOPIGELKGS 396
Db 280 LIKSQDITSYSHSKDYNPDHGYRSDSATLDEMKNQYTVOWANN--VIVHGSIGA----- 333

Qy 397 WGVQYILGQKSSALSATSEAVKQPMLLDNKVQHSYFEGVEQANW-----DNFTLRG 446
Db 334 -GVDM--QKQTTPTGTG-----YVEDGYDQRTGIVLTGLQVQGDFTFEG 375

Qy 447 GYRVEKQASIRYDKALIDRENYIKOPLDPAHROTARSFALSNWYFTPOHLSLTAS 506
Db 376 AARSDDNS-----QFGRH-----GTWQTSAGWEPFEGYRFTASYG 410

Qy 507 HQERLPSTOELVYAHGKHVATNTEFVGNKHLNKNERSNNIELALGYE-----GDRWQYNLALY 562
Db 411 TSKAPNLGQLYGF-----YGNPNLDPEKSKQWEGA--FEGLTAGVNMW--ISGY 456

Qy 563 RNRFGNYIYAQTINDRGPKSIEDSEMKLVRYNQSGADFYGAEGEIIYFKPTPRYRIGVS 622
Db 457 RNDVSLI-----DYDDHTLKYNEGKARIKGVATANFTDGLPLHT--VS 500

Qy 623 GDYVGR 629
Db 501 YDYVDAR 507

```

RESULT 13

```

YINC_ECOLI
ID YINC_ECOLI STANDARD; PRT; 700 AA.
AC P76115;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable tonB-dependent receptor yncD precursor.
GN YncD OR B1451.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
FA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
RA Riley M.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;
RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
RA Mau B.; Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000242; AAC74533.1; -.
DR PIR; F64897; F64897.
DR EcoGene; EG13774; yncD.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep Rec; 1.
DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE NEG.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;
KW Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 700 PROBABLE TONB-DEPENDENT RECEPTOR YNC_D.
FT SITE 680 700 TONB C-TERMINAL BOX.
SQ SEQUENCE 700 AA; 77260 MW; 7B3B96C6ABA48FE4 CRC64;

Query Match
Best Local Similarity 4.3%; Score 174.5; DB 1; Length 700;
Matches 166; Conservative 122; Mismatches 292; Indels 283; Gaps 41;

Qy 5 TLKPIVLSILLNTPLLAQAHETEQSVGLTETVTVGKSPRATSGLLHTSTASDKIIISGD 64
Db 5 SVRQTVLPALLVSPVFAADEQ-----TMIVSAAPQVVSEL--DTPAAVSVDGE 53

Qy 65 TLR--QKAVNLGDALDGVPGI--HASQVY-----GGASAPVIRG-----QGRRI 105
Db 54 ENRLATPRNLNLSLTGVPGLQVQRNQAQDLQLSIRGFSRSTYIGIRLYVDGIPA 113

Qy 106 KVLNHHGETGDMADFSPDHAIMVDFTALSOQVEILRGVTLTLYSSGNVAG--LVDVADGK 162
Db 114 TFPDGGQGTSN-----IDLSSVQNVVEVLRGFFSALY--GNASGVNMVTTQTGQ 160

Qy 163 IPEKMPENGVSGL-----GLRLSSGNLEKLTSGGINIGLKNFVLTGLYRKSQDYAV 217
Db 161 QPPTTEASSYYSFGSRYGLKATGATGDTGTFQGDV-----DYTE 200

Qy 218 PRYRLKRLPDSPPRFANGQRAVLGWRKRYRRTYSDRRDQVGLPAHSHEYDDCHADII 277
Db 201 -----STTRFT-----THGYRD-HSGA-- 216

Qy 278 WOKSLINKRY-----LQLYPHLLTEEDVDVNDPGLSCGFHDDDDAHAAHAKPW 328
Db 217 -QKNLANAKLVGRIDEASKLSL---IFNSVDIKADDPG---GLTKAE-----WKANPQ 263

Qy 329 DLNRKRYELRAEWKOPFPFGFEALRVHLNPNNDVHHDEKAGDAVENFNNTQONARIELRHQ 388
Db 264 APRASQYDTRTKTIKOTQAGLRYSRLSSKDDMSVMYAGERETTOYQSI:PMAPQLNPSH- 322

Qy 389 PIGRLKSGMV-----QYLQKQSSALSATSEAVKQPMLLDNKVQHSYFEGV----- 434
Db 323 -----AGGVITLQRHYOG-----IDSRWTHRGELGVPTFTTGLN 357

Qy 435 -----EQANWDNFTLEGGRVVEKOKASIRYDKALIDRENYIKOPLDPAHROTARSFA 488

```

```

Db 358 YENSMENKGYNNFLNSGMEYQKGELRDE-----RNLMWNIDPYLQWQSLSEKLS 412
QY 489 L-----SGNMYFTPOHKL-----SLTASHOERLPSQOELY-----AHGKHYA 525
Db 413 LDAGVYSVWFSDNDHYVTPGNGDSDGASHVKWLPAGSLKYAMTDANNIYLAAGRGFE 472
QY 526 TNTFE-----VGNKHLNKNRSNNIELALGYE---GDRWOYNLALYRNFGNVIYAQT 574
Db 473 TPTINELSVRADQSGMNLGLKLPSTNDTIEIGSKTRIGD-GILLSLALFQTDTHDDIIVVDS 531
QY 575 LNDGSGPKSIEDDSMKLVRYNQSG-ADPYGAEIGIYFKPTPRYRIGVSGDVVRGLKNL 633
Db 532 SSGGR-----TTYKNAGTRQGAELAWDQRFAGDFRNASWTWLDATYRS- 577
QY 634 PSLPGREDAYGNR-PFIAQDDQNAVPAPARLGFHLKASLTDRIDANLDYVRV-----FAQN 689
Db 578 -NVNCEQDCNGNRMFGIARN-----NGF-----ASIGYPEDGWYAGT 614
QY 690 KIARY-----ETPTPGHMLNLGNARYNRTRYGEWNVYKADNLLNQSIVYAHSPF 739
Db 615 E-ARYMGDIMADDENTAKAPSYTLVGLFTGYKYNYHNLTVDLFGRVDNLFDEK-YVGSVI 672
QY 740 LSDT-----PQMGSEFTGGVNV 756
Db 673 VNESNGRYEFPGRNYGVGMNI 695

RESULT 14
HPUB NEIMA
ID HPUB NEIMA STANDARD; PRT; 810 AA.
AC Q9UWA2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin-haptoglobin utilization protein B precursor.
GN HPUB OR NMA0474.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL162753; CAB83769.1; --
DR PIR; A81965; A81965.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.

```

```

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Signal; Receptor;
Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION
PROTEIN B.
FT SITE 793 810 TONB C-TERMINAL BOX.
SQ SEQUENCE 810 AA; 90570 MW; D38DE1DCA3CASA6E CRC64;

Query Match
Best Local Similarity 4.2%; Score 169.5; DB 1; Length 810;
Matches 186; Conservative 122; Mismatches 345; Indels 251; Gaps 53;

QY 7 KPIVLSILLINT-PILAAQAHETEQSVGLETTVTVGKSRPRATSGLLHTSTASDKIISGPT 65
Db 6 KPVLAALAAIAQAFPAADPAPOSAQTLEITVTGTHK-----TQKLGEEKIRKT 56
QY 66 LRQKAVNLGDALDGV---PGIHASQYG-GGASAPVIRGQTGRRIKV----- 107
Db 57 LDKLLVN--DEHDLVRYDFGISVVEGGRAGSGFTIRGVDKDRVAINVDGLAQAESRSSE 114
QY 108 -----LNHHGETGDMADFPDHAIMVDVDTALSSQOVELRGPVTLTYSSGNVAGLV 156
Db 115 AFQELFGAYGNFNAIRNTSEPNFS-----EVTITKGADSLKSGSGALGGAV 161
QY 157 DV-----ADGKIPEKMPEN-GVSGELGRLSLSGNLEKLTSGGINIGLGNFVL----- 203
Db 162 NYQTKASDYYVSEDKPYHLGIKGG-SVGKNSQKFSSTAAAGRLFGLDALLVYVRRFGKET 220
QY 204 ---HTEG-----LYRKSGDYAVPRY-----RNLKRLDPSPRPA----- 234
Db 221 KNRSTEGDIEIKNDGYVYNPTDTGGPSKYLTYVATGVARSQDPDQEWNVKSTFLKLYNPF 280
QY 235 NGQRAVLGWRKRFVRRVTVSDRRDQVGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHL 294
Db 281 NDQNR--IGW---IFEDSRDTR-----FNELSN-----LWTGTTTSAATGD-YRH- 320
QY 295 LTEDVDY-DNPGLSGCFHDDDDAHAAHNGKPMIDLRNKRYELRAEWKQPPFGEALRV 353
Db 321 --RQDVSVRRSSGVE--YKNELEHG-----PWDSLK-LRYD-----KQ-----RI 355
QY 354 HLNRNDY-----HDEKA--GDAVENF--FNQOTNARIELRHQPIGRLGKSGVQVY---L 402
Db 356 DMNTWTWDIPKNYDKRGINGEYTHSFRIHQNTAQWTDADFQK-LDFSXAVAAQYGLGG 414
QY 403 GOKSSALS-----ATSEAVKQPMLDNKNVQHSYFIQVEQANWDN-FTLEGVY 448
Db 415 GKGDNANDSYVFALYDPKILASNOAKITMLIENR-SKYKF-----ATWNAFHLGGND 468
QY 449 RVEKOKASIRYDK---ALIDRENY---YKQPLPDILGAHRQTAR-SFALSNGNMYFTPHKL 501
Db 469 RP-RLNAGIRYDKNSSSAKDDPKYTTAIRGQIPHLGSEAHAGFSYGTGDFNRFTKHLHL 527
QY 502 SLTASHOERLPSQOELYAHGKHVAINTVEVGNKHLNKNRSNNIELALGVEGDRWQYNLAL 561
Db 528 LAKYSTGFRAPTSDETLWLLFPH--PDFYLKANPNLKAFAKAKNWLGLAGSGKAGNFKLGS 585
QY 562 YRNRFNGYI---YAQTLNDGRGPKSIEDDSMKLV-----RYNOSGADFFVGAEGEYVF 611
Db 586 PKTKYRDFIELTYMGSDDDKNNPRYAPLSGDTALVSSPVWQNRSAWVGIE-----F 641
QY 612 KPT-----PR-YRIGVSGDYVRGLKNLPSLPGREDAYGNRPFIAQDDQNAF----- 657
Db 642 NGTNWLDISGLPKGLHTGLNVSYIKGK--ATQNNKGK-----TPINALSPWTVYSLGY 693
QY 658 RVPARLGFHLKASLT-----DRIDANLDYVRVPAQNKLYARETRTPGHMNLG----- 707
Db 694 DAPSKRWGINAYATRTAAKPSDTHVSNDDLNNPWPYAKHASKAYTLFDLSAYLNTGKQVT 753
QY 708 ---ANYR-RNTRYGEW-----NWTVKADNLLNQSIVYAHSSFLSDTPQMGSEFTGGV 754
Db 754 LRARAYNITNKQYTWYWSLRISIRFEGTVNRVDNKTAGIQRTIS-----PGRSYNFTI 806

```

```

QY 755 NVKF 758
Db 807 EAKF 810

RESULT 15
RHTA RHIME STANDARD; PRT; 746 AA.
AC Q923Q5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rhizobactin receptor precursor (TonB-dependent siderophore receptor
DE rhtA).
GN RHTA OR RA1265 OR SMA2414.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=21172875; PubMed=11274118;
RA Lynch D., O'Brien J., Welch T., Clarke P., Cuiv P.O., Crosa J.H.,
RA O'Connell M.;
RT "Genetic organization of the region encoding regulation, biosynthesis,
RT and transport of rhizobactin 1021, a siderophore produced by
RT Sinorhizobium meliloti."
RL J. Bacteriol. 183:2576-2585 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gozy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE RHIZOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF110737; AAD09419.1; -.
CC DR EMBL; AB007312; AAK65923.1; -.
CC DR PIR; A95420; R95420.
CC DR PIR; T46821; T46821.
CC DR InterPro; IPR000531; TonB_boxC.
CC DR Pfam; PF00593; TonB_dep_Rec; 1.
CC DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
CC DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Plasmid; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 746 RHIZOBACTIN RECEPTOR.
FT SITE 40 47 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80633 MW; 16AB44A4025D5B23 CRC64;

Query Match 4.0%; Score 161; DB 1; Length 746;
Best Local Similarity 18.9%; Pred. NO. 0.0018;
Matches 165; Conservative 124; Mismatches 306; Indels 278; Gaps 41;

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OX Enterobacteriaceae; Escherichia.
 RN NCBI_TaxID=562;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86278160; PubMed=3015941;
 RA Lundrigan M.D., Kadner R.J.;
 RA "Nucleotide sequence of the gene for the ferrienterochelin receptor
 FT PepA in Escherichia coli. Homology among outer membrane receptors
 RT that interact with TonB.";
 RL J. Biol. Chem. 261:10797-10801(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RL [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:1137-155(1996).
 RN [5]
 RP SEQUENCE OF 1-77 FROM N.A.
 RX MEDLINE=89066678; PubMed=2974033;
 RA Pettis G.S., Brickman T.J., McIntosh M.A.;
 RA "Transcriptional mapping and nucleotide sequence of the Escherichia
 RT coli fepA-fes enterobactin region. Identification of a unique
 RT iron-regulated bidirectional promoter.";
 RL J. Biol. Chem. 263:18857-18863(1988).
 RN [6]
 RP MOLECULAR ANALYSIS.
 RX MEDLINE=90354449; PubMed=2201687;
 RA Armstrong S.K., Francis C.L., McIntosh M.A.;
 RA "Molecular analysis of the Escherichia coli ferric enterobactin
 RT receptor PepA.";
 RL J. Biol. Chem. 265:14536-14543(1990).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=99101384; PubMed=9886293;
 RA Buchanan S.K., Smith B.S., Venkatramani L., Xia D., Esser L.,
 RA Palnitkar M., Chakraborty R., van der Helm D., Deisenhofer J.;
 RA "Crystal structure of the outer membrane active transporter PepA from
 RT Escherichia coli.";
 RL Nat. Struct. Biol. 6:56-63(1999).
 CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
 CC UPTAKE BY BINDING FERRIENTEROACTIN (FE-ENT), AN IRON CHELATIN
 CC SIDOPHORE THAT ALLOWS E. COLI TO EXTRACT IRON FROM THE
 CC ENVIRONMENT. PEPA ALSO ACTS AS A RECEPTOR FOR COLICINS B AND D.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M13748; AAA65994.1; -;
 DR EMBL; AB000163; AAC73685.1; -;
 DR EMBL; U82598; AAB40783.1; ALT_INIT.
 DR EMBL; D90700; BAA35225.1; -;
 DR EMBL; J04216; AAA23756.1; -;
 DR F01; F64791; QRECF0.
 DR P0B; 1FEP; 13-JAN-99.
 DR EC02DBASE; D079.0; 6TH EDITION.
 DR EcoGene; EG10293; fepA.
 DR InterPro; IPR000531; TonB boxC.
 DR Pfam; PF00593; TonB dep Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Transmembrane; Outer membrane; Iron transport; Transport; TonB box;
 KW Signal; Receptor; 3D-structure; Complete proteome.
 FT SIGNAL 1 22
 FT CHAIN 23 746 FERRIENTEROACTIN RECEPTOR.
 FT SITE 34 41 TONB BOX.
 FT SITE 729 746 TONB C-TERMINAL BOX.
 FT CONFLICT 152 152 A -> R (IN REF. 1).
 FT CONFLICT 403 403 MISSING (IN REF. 1).
 FT HELIX 40 44
 FT TURN 45 46
 FT TURN 48 49
 FT STRAND 50 54
 FT HELIX 55 60
 FT TURN 67 67
 FT HELIX 68 71
 FT TURN 72 73
 FT TURN 75 76
 FT STRAND 77 81
 FT TURN 84 85
 FT TURN 87 90
 FT STRAND 92 96
 FT TURN 97 98
 FT HELIX 101 103
 FT STRAND 104 108
 FT TURN 109 110
 FT STRAND 111 112
 FT HELIX 115 118
 FT TURN 123 124
 FT TURN 133 134
 FT HELIX 138 140
 FT STRAND 141 148
 FT TURN 149 149
 FT HELIX 150 153
 FT TURN 154 155
 FT TURN 157 158
 FT STRAND 162 168
 FT STRAND 176 186
 FT TURN 189 190
 FT STRAND 194 204
 FT STRAND 209 220
 FT TURN 225 231
 FT STRAND 235 236
 FT HELIX 237 239
 FT TURN 240 241
 FT STRAND 249 264
 FT TURN 265 266
 FT STRAND 267 281
 FT HELIX 294 298
 FT TURN 299 300
 FT STRAND 304 318
 FT TURN 320 321
 FT STRAND 324 338
 FT STRAND 360 378
 FT STRAND 382 397
 FT STRAND 426 441

```

FT STRAND 446 457
FT TURN 458 460
FT STRAND 461 473
FT TURN 476 477
FT STRAND 478 489
FT TURN 493 496
FT STRAND 498 499
FT TURN 501 503
FT STRAND 517 519
FT TURN 522 523
FT STRAND 527 540
FT TURN 541 542
FT STRAND 543 560
FT TURN 567 567
FT STRAND 569 570
FT TURN 573 597
FT STRAND 598 599
FT TURN 600 614
FT STRAND 615 617
FT TURN 627 638
FT STRAND 639 640
FT TURN 641 650
FT STRAND 653 653
FT STRAND 658 658
FT TURN 660 661
FT STRAND 664 664
FT HELIX 667 670
FT STRAND 673 673
FT STRAND 676 686
FT TURN 691 698
FT STRAND 700 701
FT TURN 705 705
FT STRAND 714 715
FT TURN 717 718
FT STRAND 722 723
FT TURN 727 728
FT STRAND 733 733
FT TURN 737 746
FT STRAND 746 746
SQ SEQUENCE 746 AA; 82107 MW; 09348AAB1C29710A CRC64;

Query Match
Best Local Similarity 4.0%; Score 160.5; DB 1; Length 746;
Matches 135; Conservative 87; Mismatches 239; Indels 257; Gaps 34;

QY 14 LLINTPL--LAQAEH-TEQSVGLTETVTVCKSPRATSGLLHTSTASDKLIISDTLRQKA 70
Db 10 LLVNLGIYGAQAEPTDTPVSHDDTIVVTAEQNLQAPGVSTITA-----DEIRKNP 62
QY 71 V--NLGDALDGVPGIHASQYGGASAPVIRGQTRRIKVLNHHGETGDMADFPDHAIMV 128
Db 63 VARDVSKIITWPGVNL--GNSTG-----GQGGNNRQI-----DIRGMPENTLIL 107
QY 129 -----DTA-----LSQVEILRGPTVLLYSNGVAGLVDA 159
Db 108 IDGKPVSSNRVQRGWRGERTRGDTSWVPPEMIEIRLGEVPAAYRNGAAGGVYNI 167
QY 160 DGK-----IPEKMPEN-----GVSGELGLRLS---SGNLEKLTSGGINI 195
Db 168 TKKGSGEWGSDWAYNAPHEKEGATKTNLSLTGPDGERSFRLYGNLDTKTQADAWDI 227
QY 196 GLGKNFVLHTEGLYRKSGYAVPRYRNKRLPDSPPRFANGQHRVILGW-----RK 246
Db 228 NQG-----HOSA---RAGTYAT-----TLPAGREGVINKDINGVVRWDPAFLQSLELE 272
QY 247 REYRR-----TYSDRDQVGLPAHSHEYDDCHADII-----WQSLNKRVLQLY 291
Db 273 AGYSROGLNLYAGDTQNTNSDSYTRSKYGDENRLYEQNYALTWNGGWDNGVTTSNNWQ-Y 331
QY 292 PHL-----LTEDVDYDNPGLSCGFHDDDDAHAAHAKGKPIDLR 331
Db 332 EHTNRSRIPEGLAGGTGEGKNEKATQDFVDIDL-----DDVNLHSEVNLPIDFILV 381
QY 332 NKRYELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFFNNTQNRATIELRHQPTG 391

```

```

Db 382 NOTLTILGTEWNNQ-----ORMKDLSSN-----TQ----- 404
QY 392 RLKSGWGVYLGQKSSALSATSSEAVKQPMMLDNKVQHYSPFFGVEQANWONFTLEGGVRVE 451
Db 405 ALTGT-----NTGCAIDGVSTDRSPY---SKAEIFSLFAENNMELDTSTI----- 447
QY 452 KOKASTRYDKALIDRENYKQPLDGLGAHROFARSFALSGNMYFTPOHKLSLTASHOERL 511
Db 448 -VTPGLRFDHHSIVGNW---SPALNISQGLGDDFTL-----KMGIARAY--KA 490
QY 512 PSTQE-----LYAHGKHV---ATNTFEVGNKHLNKRSGNNIELALGYEGDRQYNLALY 562
Db 491 PSDYQTNPNYILYSKGGCYASAGGCGYLGQNDOLKAETSINKSIGLEFKRDKGLAGVTWF 550
QY 563 RNRFGNYIYAQTILNDGRGPKSIEDDSEMKLVRYNQSGADFY-----CAEGEI 609
Db 551 RNDYRNKIEAGYVAVGQ-----NAVGTDLYQWDNVPKAVVEGLESL 592

RESULT 17
TBPI_HAEMIN STANDARD; PRT; 912 AA.
AC P44970:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable transferrin-binding protein 1 precursor.
GN TBPA OR TBPI OR HI0994.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klevavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Pine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD."
RL Science 269:496-512 (1995).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U32780; AAC22656.1; -.
DR FIR, C64107; C64107.
DR TIGR, HI0994; -.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep Rec. 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box; Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 50 57 TONB BOX.

```


FT SITE 895 912 TONE C-TERMINAL BOX
SQ SEQUENCE 912 AA; 103233 MW; 294C08991A652CC2 CRC64;

Query Match 3.9%; Score 158.5; DB 1; Length 912;
Best Local Similarity 18.9%; Pred. No. 0.0037;
Matches 194; Conservative 133; Mismatches 314; Indels 383; Gaps 51;

QY 5 TLKPI---VLSILLINTPLLAQAH---EFAQSVG-----LETVTVG-KSRP 44
DB 2 TKPYFRLSIIISCLLSYCYKAETQSTKDTKEAISSEVDQSTQSELETSIAETKVRD 61

QY 45 RATSGLHTSTADKII--SGDTRQKAVNLGDALDGPCHASQVGGGASAPV-IRGQT 101
DB 62 RKDNEV---TGLGKIKTSESISREQVNIIRDLTRYDFGIVSVQGRGASSGYSTRGMD 117

QY 102 GRIKIVL-----NHHGETGMADFPDHAIMVDLTALSOQVEILRGP 142
DB 118 RNRVALLVDGLPQTQSVVQSPVARSYSGTGAINIEYENV-----KAVEISKGG 169

QY 143 VTLVSSGNVAGLV-----DVAAG-----KIPKMPENGVSGLGLSLSSGNLE 186
DB 170 SSSEYNGALAGSVTQSKSAADILEGDKSWGTQTKNAYSKKNGFTHSLAVAGKQGGFE 229

QY 187 KLTSGGINIGLGNFVLHTEGLYKSGDYAVPYRNKLKLPDSPRRFANGQHRAVLGWRK 246
DB 230 ---GLAIYTORNSI---ETQVHKDALKGVQSVNRLIAKFPENQSAFVWNECPKG--- 278

QY 247 RFRRYSRRDQYGLPAHSHVEDDC-----HADIWQKSLN-KRYL---OLYPHLLT 296
DB 279 ---YDECIPSAKPAILLSTKKTETVSVDYTGANRIKENPMK 316

QY 297 EEDVDYDNPGLSCGFHDDDDAHAAHNGKPIWDLNRKRYELRAEMKQPP------GF 348
DB 317 YESQSFWLRG---GYHFSEQ---HYTGG---IFFTQKFDIR---DMTFPAYLSPTEKGD 365

QY 349 EALRVHLNRNDY---HHDEKAGDAVEN---FFNNOTQNAIRIELHOPIGRLKSGWQYVL 402
DB 366 LANRPVPKQDYGAYQHIEDGRGVKYGASGLYFDEHHRKQV-----GLEVI 411

QY 403 GOKSSALSATSEAV---KOPMLDNKVOH-----YSFEGVE--- 435
DB 412 YENKKGAIIDKAVLSANQONIIILDSYMRTHCSLYPNPSKNCRPTLDKPYSTRSDRVN 471

QY 436 ---QANW-----DNFT----- 443
DB 472 YKEKHNLQNLLEKKIQONWLTHQIVENLGFDFTSALQHKVLTERRVIATADISDKTG 531

QY 444 ---LEGGVRVEKQASIRYDKA---LIDRENY-----KQ 472
DB 532 KTRRNGRLREYPILYPKPKPYFAGEDHCNYSQSSNRYRDCVKRLIKGKNYFYAARNNALG 591

QY 473 PLPDLG-----AHRQTA-----RSFALSGNWFYFPQHKLSIT--ASHQERLPST 514
DB 592 KYVDLGLGIRYVSRTKANESTLSVKFKPNFVNTGLIVKPTETWLDLSVRLSTGFRNPSF 651

QY 515 QELYAHGKHVATNTFFVGNKHLNKNERSNNIELALGYEGDRWQYNLALYRNFQNV1-YAQ 573
DB 652 AEMY--GWRYGGKNDVEYVYKPKPEFSRQEFGLALKGDFGNTIEISHFSNAYNLIAFAE 709

QY 574 TLN---DGRGPKSIEDSEMKLVRYN-QSGADPYGAEIYF----- 611
DB 710 ELKNGTGRKNGYHNAQNAKLGVNITAQLDPNGLWKLRIPIYGYWATFAYNRVKVKDQKI 769

QY 612 ---KPTPRYRIGVSGDYV-----RGLKNLPSLPQREDAYGN 645
DB 770 NAGLASVSVLFAIQPS-RYIIGLVYDHLISNTWGNVATFTQSKAKSQNELLGK-ALGN 827

QY 646 RPFIAQDQONAPVPAARLGFHLKASLITRIDANLDYRVFAONKLARVETRPFQGHMLN 705
DB 828 N---SRDVKSTRKLTRA---WHI-----LDVSGYWMANK-----NIMLR 860

QY 706 LGA-----NTR-----RNRTRYGEWNNYKADNLLNQSVIAHSSFLSDTFQMGSRSTGGV 754

Db 861 LGIYNLFNRYVTWVAVRQTAQAVNQH---QNVGNTRYAAS-----GRNYTLTL 908

QY 755 NVKF 758
:|||

Db 909 EMKF 912

RESULT 18
PFEA PSEAE
ID PFEA PSEAE STANDARD; PRT; 746 AA.
AC Q05038;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric enterobactin receptor precursor.
GN PFEA OR PA2688.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K407;
RX MEDLINE=931123148; PubMed=8419284;
RA Dean C.R., Poole K.;
RT "Cloning and characterization of the ferric enterobactin receptor
gene (pfeA) of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:317-324 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
ENTEROBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY IRON AND ENTEROBACTIN.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M38033; AAA25928.1; -;
CC EMBL; A5004697; AAG06076.1; -;
CC PIR; A40636; A40636.
CC HSP; P05825; 1FEP.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rcg; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 25 POTENTIAL
FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR.
FT SITE 39 44 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80967 MW; ADD4FB0CB23251 CRC64;
Query Match 3.9%; Score 155.5; DB 1; Length 746;
Best Local Similarity 19.3%; Pred. No. 0.0045;


```

QY 270 -----DDCHAD-----IIWQSLINKRYLQLYPHLITEED----- 299
Db 364 LGKYSQGNKAERLRFVQEGSTLQIGYGTGVYDERHTKRYVEYVYHNADKDTWADYA 423
QY 300 -VDYDNFGLS-----CGFHDDDAHAHNGKPMIDLRNKR--YE-----LRAEWK 342
Db 424 RLSVDRQGIIDNLFQTHCS-HGSDKNCPCPDGKPYSFYKSDRMIEYBSRNLFOAVFK 482
QY 343 QPPF-----GFELRVLHNRNDYH--HDEKAGDAV-----ENFFNQTON----- 380
Db 483 KAFDTAKIRHNLINLGYDRFKSLSHSDYVQLQNAQVAYDLITPKPPFPNGSKDNPRV 542
QY 381 --ARIELRHOPIGLKGSGWQVYLGQKSSALSATSEAVKQPMLLDNKVQVSPFGEQAN 438
Db 543 SIGITVTNTSPICRFGNN-----TYDCTPRNTGNGGYAAV 579
QY 439 WDNFTLEGGVRVERQKASIRYDKALIDRENYKQPLDPAHROTARSFALSGN----- 492
Db 580 QDNVRLG---RWADVAGIRYDY----RSTHSEDKSVSTGTHRN-----LSWAGVVLK 626
QY 493 ---WYFTPOHKLSTASHQERLPSTQELYAHGKHVANTFEVGNKHLNKRNNIELALG 549
Db 627 PFTWM-----DLTYRSTGTFRLPSFAEMYGWRAGESLKTLD-----LKPEKSFNREAGIV 676
QY 550 YEGDRWQYNLALYRNFQNYI---YAQTLNDGR---GPKSIEDDSEMKLVRYNQSG-AD 601
Db 677 FKDGFGNLEASYFNNAIRDLIAFGYETRTQNGTSASGDPGYRNAQNRAGINILKID 736
QY 602 FYGAEGBIYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFAQ--DDQNAPRV 659
Db 737 WHGVWGLG---PDGLYS---TLAYNRKVKD-----ADIRADRTFTVSYLFDVAPQSR 783
QY 660 PAARLGF-----HLKASLTDR-----IDANLDYRVRFAQNKLYARTYTRT 698
Db 784 YVLGLGYDHPDGIMGWINTMETYTSKAKSDELLGSCALLNGANAKAASRRTRPWWYTV 843
QY 699 PG-----HMLNLGANYRNTRYGEM-NWYVKADNLNQ--SVYAHSSFLSDTPQWGRS 749
Db 844 SGYINIKHLTLRAGVYNLLNRYVTWENVRQTAGGAVNHQVGNRYNAAP-----GRN 899
QY 750 FTGGVNVKF 758
Db 900 YTFSEMKF 908

RESULT 20
ID IUTA_ECOLI STANDARD; PRT; 732 AA.
AC P14542;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ferric aerobactin receptor precursor (Cloacin receptor).
GN IUTA.
OS Escherichia coli.
OG Plasmid IncFI ColV3-K30.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F344;
RA Krone W.J.A., Stegehuus F., Koningsstein G., van Doorn C.,
RA Roosendaal B., de Graaf F.K., Oudega B.;
RT "Characterization of the pColV-K30 encoded cloacin DF13/aerobactin
RT outer membrane receptor protein of Escherichia coli; isolation and
RT purification of the protein and analysis of its nucleotide sequence
RT and primary structure.";
RL FEMS Microbiol. Lett. 26:153-161(1985).
RN [2]
RP REVISIONS.
RA Oudega B.;
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.

```

```

CC --!- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.
CC --!- SUBCELLULAR LOCATION: Outer membrane.
CC --!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05874; CAA29297.1; --
DR EMBL; X05874; CAA29298.1; ALT INIT.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Plasmid;
KW Receptor.
FT SIGNAL 1 25
FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.
FT SITE 31 38 TONB_BOX
FT SITE 715 732 TONB_C-TERMINAL_BOX
SQ SEQUENCE 732 AA; 81014 MW; 0C23879C0B27AE2B CRC64;

Query Match 3.8%; Score 154; DB 1; Length 732;
Best Local Similarity 19.9%; Pred. No. 0.0055;
Matches 168; Conservative 109; Mismatches 301; Indels 266; Gaps 45;

QY 6 LKPIVLSILLINPTLLAQAHETEVSGLQETVVVVKSRPRA-----TSGLLHTSTASDKII 61
Db 12 LNPULLTMM---APAVAQQTDDTFV---VSANRSNRTVAEMAQTWVIENAELEQQIQ 64
QY 62 SGDTLRQKAVNLGDLQVPGIHAS-----QYGGGA-SAPVIRQOTGRIRKVLNHHGTG 115
Db 65 GKELKDALLAQ-----IFGLVSSRSRTYGNVGRPLVLVDGVRLLNSTRSDTSRQL 118
QY 116 DMAD-FSPDHAIMVDTALSOQVEILRGFVITLLYSSGNVAGLDVADGK-IPEKMPENGVS 173
Db 119 DSIDPFNMWH-----IEVIFG-ATSLYGGSTGGLINIVTKGQPTWMEFEAG 166
QY 174 GEIGLRSLSSGNLEKL-----TSGGINIGLGNFVLHTEG-----LYRKSQGYAVPRYRNKRL 226
Db 167 TKSGFSSSKDHERIAGAVSGNE-----HISGRLSVAYQKFGW----- 206
QY 227 PDSPRRFANGQRAVL-----GWRKFRYRTYSDRDQVGLPAHSHEYDDCHADLIWOK 280
Db 207 -----FDGNGDATLLDNTQTGLQ-----YSDRLDIMGT-----GTLNIDESRQL 245
QY 281 SLINKRYLQLYPHLLTBEVDYDNPGLSCGFHDDDDAHAAHNGKPMID--LRNKRYELR 338
Db 246 QLITQYY-----KSQGDYD---GLNLG---KGFSAIRGTSTPFPVSNGLNSDR---- 287
QY 339 AEWKQPPFPGEALRVHLNRND-----YHDE-----KAGDAVENFNN 376
Db 288 -----IPGTDGHLISLQYSDSAFLGQELVGQVYRDESLRFYPPFTVNAANKQVTAFFSS 341
QY 377 QTO---NARIELRHOPIGRLKSGWQVYLGQKSSA-----LSATSEAVKQPMLLD 423
Db 342 QQTDQYGMKLTLSNKPMDGQWQITWGLDADHERFTSNQMFDFDLAQASAGSLNNKIYTT 401
QY 424 NKVQHYVFFQVE---QANWD---NFTLEGGVRVEKQKASIRYDKALIDRENYKQPLD 477
Db 402 GRYPYSDITNLAAFLQSGYDINNLTFLNGGVRYQ-----YTNKIDDFIGYAAQ-RQI 453
QY 478 GARRQTA-----RSFALSGN--WYFTPOHKLSTASHQERLPSTQELYAHGKHVA 525
Db 454 GAGKATSADAFWRLSLRHLFNLNAGLMLHTEPQQAWLNFESQGLELPDPGKYVGRGIYGA 513
QY 526 -----TNTFEVGNKHLNKRNNIELALGVEGDRWQYNLALYRNFQNYIAQ---T 574
Db 514 AVNGHLPLTKSVNVSDSKLEGVQVDSYELGWRRTGNLRTQIAAYYSISDKSVVANKDLT 573

```

Qy 575 LNDGRGPKSIEDSEMKLVRYNQSGADFYGAGEI-YFKPTPRYRIG-----VS 622
 Db 574 I-----SVDDKRR-----IYGVGAVDYLIPDWDSTGVNENVLKTESKVN 615
 Qy 623 GDYVGRGLKNLPSLPGRDAY-----GNRPPIAQDDQ-----NA 656
 Db 616 GTWQYDYKVT--ASPKATAYIGNAPDWSLRVQSTTFDVSDAQGYKVDGYTTVDLLGS 673
 Qy 657 PRVPAARLGFHLKASLTDRIIDANLDYRVFAQNKLARLYETRTPGHMLNL----- 706
 Db 674 YQLPVGTLFSFIE-NLFDK-----DYTWGQRAPLY--SPGPGASLDYKGRGRTF 724
 Qy 707 GANY 710
 Db 725 GLNY 728

RESULT 21
 FOXA_YEREN STANDARD; PRT; 710 AA.
 AC Q01674;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Ferrioxamine receptor precursor.
 GN FOXA.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
 RX MEDLINE=92349959; PubMed=1640832;
 RA Baeumler A.J., Hantke K.;
 RT "Ferrioxamine uptake in Yersinia enterocolitica: characterization of
 the receptor protein FoxA".
 RL Mol. Microbiol. 6:1309-1321(1992).
 RN [2]
 RP REVISIONS.
 RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
 RA Baeumler A.J.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: FERROXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
 TONB PROTEIN.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X60447; CAA42975.1; -.
 DR HSSP; P06971; IQQJ.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 KW Transport; TonB box.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 710 FERROXAMINE RECEPTOR.
 FT SITE 28 35 TONB_BOX.
 FT SITE 693 710 TONB_C-TERMINAL_BOX.
 FT TRANSMEM 29 37 POTENTIAL.
 FT TRANSMEM 65 73 POTENTIAL.
 FT TRANSMEM 91 99 POTENTIAL.
 FT TRANSMEM 106 114 POTENTIAL.

FT TRANSMEM 137 145 POTENTIAL.
 FT TRANSMEM 152 160 POTENTIAL.
 FT TRANSMEM 180 188 POTENTIAL.
 FT TRANSMEM 194 202 POTENTIAL.
 FT TRANSMEM 208 216 POTENTIAL.
 FT TRANSMEM 259 267 POTENTIAL.
 FT TRANSMEM 271 279 POTENTIAL.
 FT TRANSMEM 293 301 POTENTIAL.
 FT TRANSMEM 309 317 POTENTIAL.
 FT TRANSMEM 353 361 POTENTIAL.
 FT TRANSMEM 370 378 POTENTIAL.
 FT TRANSMEM 427 435 POTENTIAL.
 FT TRANSMEM 443 451 POTENTIAL.
 FT TRANSMEM 476 484 POTENTIAL.
 FT TRANSMEM 491 499 POTENTIAL.
 FT TRANSMEM 517 525 POTENTIAL.
 FT TRANSMEM 531 539 POTENTIAL.
 FT TRANSMEM 555 563 POTENTIAL.
 FT TRANSMEM 567 575 POTENTIAL.
 FT TRANSMEM 579 587 POTENTIAL.
 FT TRANSMEM 610 618 POTENTIAL.
 FT TRANSMEM 624 632 POTENTIAL.
 FT TRANSMEM 649 657 POTENTIAL.
 FT TRANSMEM 671 679 POTENTIAL.
 FT TRANSMEM 684 692 POTENTIAL.
 FT TRANSMEM 702 710 POTENTIAL.
 SQ SEQUENCE 710 AA; 78382 MW; 82EB6EC1546900B8 CRC64;

Query Match 3.8%; Score 151.5; DB 1; Length 710;
 Best Local Similarity 17.9%; Pred. No. 0.0079;
 Matches 144; Conservative 100; Mismatches 250; Indels 309; Gaps 36;

Qy 10 VLSILLINTPLLAQAHETEQSVGLETTVTVGKSRPRATSGLIHT-----STASDK--IISG 63
 Db 12 ILCSLAMPFIPLASIAADDT-----IEVTAKAGHEADLPTSGYTATTTKGATKTDQPLILTA 66
 Qy 64 DTLR-----QKAVNLGADLGVPGCIHSAIOYGGAS---APVIRGQTGRRIKVLNHH 111
 Db 67 QSVSVVTRQQMDQNVATVNOALNVTYFTG-FSGGATRYDVTVALRG-----FHG 116
 Qy 112 GETGD-----MADFSFDHAIMVDLTALSOQVEILRGFVTLTYSSGNVAGLVADVADGKI 163
 Db 117 GDVNTFLDGLRLSLDGGSYNVQLVDPWFLERIDVIRKGPSSALYQGSIPGGVV-----MM 171
 Qy 164 PERKENGVSIGELGLRLSSGNLE-KLTSGGINIGLGNKFNVLHTEGLYRKSGDYAVPRYRN 222
 Db 172 TSKRPQ--FTSEGHFLTAGNNNTQVAAFDVTDAISEHMAFLTGTITNSD----- 220
 Qy 223 LKRLPDSRPRPANGQRAVLGWRKRFYRTYSDRRDQGLPAHSHEYDDCHA--DIIWQ 279
 Db 221 -----TMYDHOEREYAIAFSLMW 240
 Qy 280 ----KSLINKRYLQLYP-----HLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKPIDLR 331
 Db 241 PDENTSLLLRANLQKDPGGYHSAVPADGSIYGQKLSRGFFDGESNH----- 287
 Qy 332 NKRYELRAWKQPPFGFEALRVHLNRNDYHDEKAGDAVENFFNNQT---QVARIELRHQ 388
 Db 288 ----NVFKRWQIY-----SYEFSHKPDD-VMSFRQNASYTHSNLTQLEQVYQ 329
 Qy 389 PIGRLKSGMV-----QYLQCKSSALSA-----TSEAVKQPMLLDNKVKQHSF 431
 Db 330 -----GGWNSDRITLMNRYISGEDSSLNFAFVNDQLEADLEATAAVKHKVLL----- 374
 Qy 432 FGVEQANWDNFTLEGGVRVEKOKASIRYDKALIDRENYVKOPL-PDLGAH----- 480
 Db 375 -----GVDFOKFRNLRSDSA-----YATPLNPYTVSGSGSTLYSDYL 412
 Qy 481 -----ROTARFALSG-----NWY--FTPQHKLSLTASHOERLPSTOEL--- 517
 Db 413 LTTPGINTSYLSRRYEQSGVYLQDEMTLDNHLNLSGRYDRMKTENINNTANSTDTDN 472
 Qy 518 YAHGKHVATNTFEVG-----NKHINKERSNNIELALGYE--GDRW 555

```

Db 473 HASGRASLLYSFDSGISPIYSVYSQAITSLSFDDAQOKLLKPMWTEQYEVGVIIYQPPGSTS 532
Qy 556 QYNLALY---RNRFGNYIYAQTLDNGRGPXSIEDDSEMKLVRYNQSGADFYGAEGEYIFK 612
Db 533 LYSAAVLDLTQNDVANRAVPATYVYPAG-----KVNSQL-----ELEAR 572
Qy 613 PTPRYGVSGDYVGRGLKNLPSLPGREDAYGNRPFA----- 650
Db 573 SQISDRUSVIAGTYNRYVKFDKIDGND---GNTPLVLAASNWASLWQAYEAGYGINVGAG 629
Qy 651 -----QDDQNAAPRVPAAALRG 665
Db 630 IRYIGKQWADDANTLRVPSYTLG 652

RESULT 22
TB11_NEIMO STANDARD; PRT; 915 AA.
AC Q01956;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TB11.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=923394880; PubMed=1325963;
RA Cornelissen C.N., Biswas G.D., Tsai J., Paruchuri D.K.,
RA Thompson S.A., Spurling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
RT utilization and is homologous to TonB-dependent outer membrane
RT receptors.";
RL J. Bacteriol. 174:5788-5797 (1992).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M96731; AAA25503.1; -.
DR F.R.; A43335; A43335.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 898 915 TONB C-TERMINAL BOX.
SQ SEQUENCE 915 AA; 102213 MW; 697CF74B1010422F CRC64;

```

```

Query Match 3.7%; Score 150.5; DB 1; Length 915;
Best Local Similarity 18.4%; Pred. No. 0.013;
Matches 186; Conservative 132; Mismatches 344; Indels 349; Gaps 48;

Qy 1 MAQTLKPIVLSILLITPLLAQHETE----QSVGLTETVVGKS-----RPRATSGLLH 52
Db 1 MQQHLFLNLICLSLMTALPAYAENVQAOQAEKQLDITQVKAQKQKTRDRNEVTGL-- 58

```

```

Qy 53 TSTASDKII--SGDTL--ROKAVNLGDALDVGPGIHAHQYGGGASAPV--IRGQTGRRIKV-- 107
Db 59 -----GKLVKTADTLTSKEQVLDIRDLTRYDPCIAVVEQGRGASSGYSIRGDKNRVSLTV 113
Qy 108 -----LNHHGETGDMADFPDPAHIMVDLTALSOQVEILLRGPVTLILYSSG 150
Db 114 DGLAQIQSYTAQAALGGTRTAGSSGAINEIEYENV-----KAVEISKGSNSVEQSG 165
Qy 151 NVAGLV-----DVADGKIPKMPENGSGELGRLRLSSGNLEKLTSGGINIGLKFNVL-- 203
Db 166 ALAGSAVAFQTKTADDVI-----GEGRWGIQ-----SKTAYSGKRLTQSTALAGR 212
Qy 204 --HTEGLYRKSGDYA-----VPRYNLKRLLPDSR-----RPAQCQH-- 238
Db 213 IGGAEALLIIRTRHAGEIRAHEAAGRGVQSFNRLAPVDDGSKYAFIVEEBCNGGHEKC 272
Qy 239 -----RAVLG-----W-----RKFY-----R 250
Db 273 KANPKQVVGEDKQTKQTVSTRDYTGPNRFLADPLSYESRSWLFPRGFRFENKRYIGGILE 332
Qy 251 RTYS--DRRQ-----YGLPAHSHEYD-----D 271
Db 333 RTQQTFTDRDMTVPAFLTKAVFDANQKQAGSLRGNGKYAGNHKGGLFTSGENNAVPAGE 392
Qy 272 CHADIIWQKSLINKRYLQLYPHLLTEED-----VDYDNPGLSCGFH-----DDDD 316
Db 393 YGTGVFYDETHTKRGYLEVYVYTNADKDTWADYARLSYDRQIGLDNHFQOHTHCADSGD 452
Qy 317 AHAHAHNGKPMIDLRNKRY-----ELRASWKOPFP-----GPEALRVHLNR 357
Db 453 KYCPASADKPFYSKSRVYGVESHKLLQAAFKKSFDTAKIRHNLVSNVGLDRFQSNLRH 512
Qy 358 NDYHHDEKAGDAVENFNNQTONARIELRHQPIGRLKGSWGVQYLGQKSSALSASEAVK 417
Db 513 QDYVY-----QSANRAYSLKTPQNNNGKTSPPNGREKNPYV-----SIGRCNVYTR 559
Qy 418 QPMLLDN-----KVQHYSPFGVEQANWNTFLEGGRVEKQKASIRVDKALIDREN 468
Db 560 QICLFGNNTYDCTPFRSINGKSYAAVR---DNVRLG---RWADVGAGLRVD----- 605
Qy 469 YYKQPLDPLGA--HRQTARSPALSGNWTFTPOHKLSLT--ASHQBELPSTOELYAHGKHA 525
Db 606 -YRTHSDDGVSVTGTRHLSWNAGIVLKPADWLDLTYRTSTGFPPLSPFAEMYG----- 658
Qy 526 TMTPEVGNK-----HLNKNRSNNIELALGYEGDRMQYNLALYRNRFNGYI---YAQTLDNG 578
Db 659 ---WRSGDKIKAVKIDPEKSPENKAGIVFKGDFGNLRASWFNNVARDLIVRGYEAQIKDG 715
Qy 579 R-----GPKSTIEDDSEM-----KL-----VRYNOSGADF 602
Db 716 KEQVKGMPAYLNAQSARITGINILGKIDWNGVDKLPFGWYSTTFAYNRVVRVDIKKRADR 775
Qy 603 YGAEGEIY--FKPT-----PRYRIGVSGDYVVRGRLLKNLPSLPQREDAYGNRPFA 650
Db 776 TDIQSHLFDALQPSRYVVGSDYDQPEGKVGWNGMLTYSKAKEITEL-----LGSRALIN 829
Qy 651 QDDQNAAPRVPAAALRGFHLKASLTDRIDANLDYRVPFAQNKRLARYETRTPGHHMLMGANY 710
Db 830 GNSRNTKATARRTRPWI-----VDVS--GYIV-----KKHFTLRAGVYN 868
Qy 711 RNRTRYGEW--NYYVKAIDLNQ--SVVAHSSFLSDTPQMGSRSPFTGCVNVKPF 756
Db 869 LLNHRVYTWENVROTAAGAVNQHNKGVYNYAAP-----GRNYTFSLEMKF 915

RESULT 23
TB11_NEIMB STANDARD; PRT; 911 AA.
AC Q09056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.

```

GN TBPI.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=CCUG 37608 / M982 / Serogroup B / Serotype 9;
 RX MEDLINE=93345825; PubMed=8344530;
 RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
 RA Jacobs E., Schryvers A.B.;
 RT "Cloning and characterization of Neisseria meningitidis genes
 encoding the transferrin-binding proteins Tbp1 and Tbp2.";
 RL Gene 130:73-80(1993).
 CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
 CC TRANSFERRIN UTILIZATION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: By iron starvation.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z15130; CAAY78833.1; -;
 DR PIR; JN0821; JN0821.
 DR InterPro; IPR00531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Receptor; Signal; TonB box.
 FT SIGNAL 1 24
 FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.
 FT SITE 38 45 TONB BOX.
 FT SITE 894 911 TONB C-TERMINAL BOX.
 SQ SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773B6 CRC64;
 Query Match 3.6%; Score 144.5; DB 1; Length 911;
 Best Local Similarity 19.0%; Pred. No. 0.035;
 Matches 191; Conservative 135; Mismatches 337; Indels 343; Gaps 52;
 QY 1 MAQTLLKPIVLSILLINTPLLAQHETE-----QSVGLETVTVVGKS-----RPRATSGLLH 52
 Db 1 MQQOHLFRNLCLSLMTALPAVAENVQAQAEKQDITQVAKKQKTRDNEVTGL-- 58
 QY 53 TSTASDKII-SGDTL-RQKAVNLGDALDGVPGTHASQYGGASAPV-IRGQTRRIKV-- 107
 Db 59 -----GKLVKATDLSKEQVLDIRDITRDYDPGIAVVEQGRGASSYIRGMDKNRVSITV 113
 QY 108 -----LNHHGETGDMADFPDPAHIMVDTALSSQVVEILRGVPVTLVSSG 150
 Db 114 DGLAQISYTAQAALGTRTAGSSGAINETEVENV-----XAVEISKGSNVSQGG 165
 QY 151 NVAGLV-----DVADGKIPEK-----MPENGVSSELGLRLSSGNLEKLTSGINIGLQNF 201
 Db 166 ALAGSVAFOTKTADDVIGEGROWIGIQSKTAYSGK-----NRGLTQSIALAG-RIGGAEL 219
 QY 202 VLHTEGLYRKSGDY-----AVPRYNLKRLPDS----- 229
 Db 220 LIHTG--RRAGEIRAHEDAGRGVQSFNRLVPVEDSSEVAYFIVECEGKNYETCKSKP 276
 QY 230 -----PRRF-----ANGQRAVL-----GMR-----KRFY-----RR 251
 Db 277 KKDVGCKDERQTVSTRDYTGNEFLADPLSYESRSLFRPGFFENKRRHYIGILEHTQQ 336
 QY 252 TVSDRRD-----QVG-IP-----AHSHEYDDCHAD----- 275
 Db 337 TF-DTRDMTVPAFLTKAVFDANSKQAGSLPGNGKYAGNHKYGGLFTNGENGALVGAEGYT 395

QY 276 -IIWQSLINKRYLQLYPHLLTEED-----VDYDNPGLSCGFH-----DDDAHA 319
 Db 396 GVFDYDEHTKSRVGLBYVTNADKOTWADYARLSYDRQIGLNDHFQOOTHCSADGSKYC 455
 QY 320 HAHNGKPDIDLNKRY-----ELRAEWKQPPP-----GFEALRVHLNRNDY 360
 Db 456 RPSADKPFYSYKSDRVYIGESHRLLOAPFKGFDATAKIRHNLNVLGFORFDSNLRHODY 515
 QY 361 HDDEKAGDAVENFFNNQTONARIELRHQPIG-RLKGSW-----GVQYLQ-----KSSAL 409
 Db 516 YY-QHANRAYSSKTPPKTAN-----PNGDKSKFYWISIGGNVVTGQICLFGNNTYT 566
 QY 410 SATSEAVKQPMLLDNKVQHSYFFGVEQANWDNFTLEGVVRVEKQKASIRYDKALIDRENY 469
 Db 567 DCTPSI-----NGKSYIA-----AVRDNVRLG--RWADVAGLRYD----- 601
 QY 470 YKQPLDPLCA-HRQTARSPALSNGWYFTPOHKLST--ASHQRLPSTOELVYANGKHVAT 526
 Db 602 YRSTHSDGVSSTGTHRTLSNAGIVLKPADMLDLYTSTGFRLPFAEYMGWRSGVQS 661
 QY 527 NTFEVGNKHLNERSNIELALGYEDRWQYNLALYRNFQNYIY---AQTNL---DGR 579
 Db 662 KAVKI-----DPEKSFNKEAGIVPKGDFGNLEASWNNAYRDLIVRGYEAQIKNGKEAK 716
 QY 580 GPKSIDDDSEMKLVRYNQSG-----ADFYGAEG 607
 Db 717 GDPAYLNAQASARITGINILGKIDWNGVDKPEGWYSTFAYNVRHVVDIKKRAEDTIQS 776
 QY 608 EIFYKPTP-RYRIGVSGDYVRG-----RLKNLPSLPQREDAYGNRPPIAODQN 655
 Db 777 HLFDAIQPSRYVYVGLGYDQPEGKWGVNGMLTYSKAKEITEL-----LGSALLNGSRN 830
 QY 656 APRVPAARLGFHLKASLTDRIANLDYRVFAQNKCLARYETRTPGHMLNLGANYRNRTR 715
 Db 831 TKATARTRPWI-----VDVS-CYITI-----KKFTLRAGYNNLLNR 869
 QY 716 YGEW-NWYKADNLLNQ--SVYAHSSFLSDTPQMGSRFTGGVNVKF 758
 Db 870 YVTWENVROTAGAVNQHNKNGVYNYAAP-----GRNYTFSLEMKF 911
 RESULT 24
 FHUE_ECOLI
 IN FHUE_ECOLI STANDARD; PRT; 729 AA.
 AC P16869; P77292;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE FhuE receptor precursor (Outer-membrane receptor for Fe(III)-coprogen,
 DE Fe(III)-ferrioxamine B and Fe(III)-rhodotulic acid).
 GN FHUE OR B1102.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=90286919; PubMed=2162465;
 RA Sauer U., Hantke K., Braun V.;
 RT "Sequence of the fhuE outer-membrane receptor gene of Escherichia
 RT coli K12 and properties of mutants.";
 RL Mol. Microbiol. 4:427-437(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X69214; CMA49148.1; --
 DR EMBL; AE002504; AAF41895.1; --
 DR PIR; G81070; G81070.
 DR TIGR; NMB1540; --

DR InterPro: IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Reg; 1
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.

FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 943 LACTOFERRIN BINDING PROTEIN A.
 FT SITE 826 943 TONB C-TERMINAL BOX.
 FT CONFLICT 8 8
 FT CONFLICT 16 18 IAT -> Q (IN REF. 1).
 FT CONFLICT 22 22 A -> S (IN REF. 1).
 FT CONFLICT 26 31 QAGGAT -> NPETA (IN REF. 1).
 FT CONFLICT 43 43 I -> V (IN REF. 1).
 FT CONFLICT 64 64 V -> A (IN REF. 1).
 FT CONFLICT 233 233 R -> H (IN REF. 1).
 FT CONFLICT 243 243 E -> A (IN REF. 1).
 FT CONFLICT 247 247 D -> N (IN REF. 1).
 FT CONFLICT 257 269 DIKRTREPPFSV -> GIKKPSGGGEYFLA (IN REF. 1).

FT CONFLICT 273 275 RES -> SEL (IN REF. 1).
 FT CONFLICT 281 281 L -> V (IN REF. 1).
 FT CONFLICT 284 286 YGK -> NGN (IN REF. 1).
 FT CONFLICT 313 313 Q -> M (IN REF. 1).
 FT CONFLICT 389 389 E -> K (IN REF. 1).
 FT CONFLICT 420 423 KNLV -> OKLI (IN REF. 1).
 FT CONFLICT 455 455 A -> K (IN REF. 1).
 FT CONFLICT 546 546 K -> N (IN REF. 1).
 FT CONFLICT 564 572 STGDEENNQ -> YSDYTDKG (IN REF. 1).
 FT CONFLICT 658 658 L -> V (IN REF. 1).
 FT CONFLICT 667 667 V -> L (IN REF. 1).
 SQ SEQUENCE 943 AA; 105680 MW; BD569ECACFC01A84 CRC64;

Query Match 3.4%; Score 138.5; DB 1; Length 943;
 Best Local Similarity 19.4%; Pred. No. 0.096;
 Matches 200; Conservative 126; Mismatches 328; Indels 379; Gaps 54;

QY 8 PIVLSILLINTPLAAQA-----HETEQSVGLETVTV---VGKSRPRATSGLLHTSTA 56
 Db 8 PLTLTALAIAATAPAYAAQAGGATPDAAQTQSLKEITVRAAKVGR-REKEATGL----- 60
 QY 57 SDKII-SGDTL-RQAVNLGDALDGVFGIHAQYGGGASAPV-IRGQTGRRIKV----- 107
 Db 61 -GKIVKTSITLKEQVLGIRDLTRYDGVAVVEQNGASGGYSIRGVDKRVAIVSDGVA 119
 QY 108 -----LNHHGETGDMADFPDRAIMVDTLSQVEILRGVILLVSSGNVAGLVD 157
 Db 120 QIOAFTVQSGLSGYGGGGGAINIEFYENIST-----VEIDKAGSGDHGSGALGAVA 174
 QY 158 -----VADGR---IPEKWPENGVSGE-----LGLRLSSGNLEK----- 188
 Db 175 FRTKEADLISDGKSWGIGQAKTAVGSKNRQPMKSLGAGFKDGEGLLITERGQRETRP 234
 QY 189 ---TSGGINTGLGK-NFVLHTEGLYRKS-----GDY----- 215
 Db 235 HGDIAADVGEYIDRLDAFQTYDIKRTREFFSVGEGRSKPVAKLAGYKYLNNQINLR 294
 QY 216 -----AVPRYRLN-----KRLDPSRRFANGQRAVLGV 244
 Db 295 WVKERIEQNQPLSABEEAQVREQAQRHENLSAQAYTGGGRILPD-PMDYRSGSWLAKIGY 353

QY 245 R---KRFVRTYSRRDOYGLPAHSHEYDDCHADIIWQKSLINKKYLQYLPHLLFEEDVD 301
 Db 354 RFGGRHYVGGVFEDTKQYDI-----RDMTEKQYVG-----TDEAEK 390
 QY 302 YDNFGLSCGFHDDDDAHAAH--NGKPW-----IDLNRKEVELRA 339
 Db 391 FRD---KSGVYDGDGDFRDGLYFVNIEEWKGDKNLVRIGLKYKSTKFIDEHRRRML 447
 QY 340 EWQFFFGFEALRVHLNRDHYHDEKAGDAVENFNN--QTQNARIEL----- 385
 Db 448 LYRY-----ENEAYSNDWADKAVLSPDKQGVATDNTLKLNCVYPVPAVDKSC 494
 QY 386 -----RHQPIG-----RLKSGWGVQY--LQOKSSALSATSEAVKQ 418
 Db 495 RASADKPYSDSSDRFHYRQHNVLNASFEKSLKNKTKHLLTGLFGYDA---SKAISR 550
 QY 419 PMLLDNKKVQHYV--FFGVQOANDNFTL---EGGV-----RYEKQKASIR--- 458
 Db 551 PEQLSHNAARISESTGFDENNQDKYLLGKPEVGVSGVYIETURSKCVPRKINGSNIH 610
 QY 459 -----YDKAL---IDRENYKQPLDGLGAHRQTARSFALSG---NW---Y 494
 Db 611 ISLNDREFSIGKYVDFSLGGYDRKNF-----TTSEELVRSRGYVDRSWNSGIL 658
 QY 495 FTPOHKLST--ASHQERLPSTQELYAHGKIVATNTFVG--NKHINKERNNIELALGY 550
 Db 659 FKPNRHFSVSYRASGFRTPSFQELFGIDY---HDYPKQWRPALKSEKAAANREIGLQW 715
 QY 551 EGRWQYNLALYRNRFNGYIYVAQTLNDGRGPKSIEDOSEMKLVRYNQS---ADFY 603
 Db 716 KGDFGLFLEISFRNRYTDMIAVA-----DHKTCLP--NQAGQLTEIDIRDY 760
 QY 604 GAE-----GEIYFKPTPRYRIGVSGDYVRGLKNLPSLPQREDAVGNRPPIAQDD 653
 Db 761 NAQWMSLQGVNLIKIDWN-----GVYKLPGLYTLTAYNRKPKSVNRPGLSLRS 813
 QY 654 QNAPRVPAAR--LGF-----HLKASLTDRIANDLYYRVAQNKLAARYETRTPGHH 702
 Db 814 YALDAVQPSRYVLGFGYQDPGKNGANIMLTYSKGNPDELAAYLAGDQ-KRYSTKRASS 872
 QY 703 -----MLNL-----GANRY-RNTRYGEW-NWYVKADNLLNQSVYAHSSFLSDTPQ 745
 Db 873 WSTADVSAYLMLKRLTLRAAIYGNVRYVTWESLQTAESTANR--HGGDSNRYGYAA 930
 QY 746 MGRSFTGGVNVKF 758
 Db 931 PGRNFSLALEMKF 943

RESULT 27

FPTA PSEAE

ID FPTA PSEAE STANDARD; PRT; 720 AA.

AC P42512;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Fe(III)-pyochelin receptor precursor.

GN FPTA OR P44221.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-51.

RC STRAIN=PAO / IA602;

RX MEDLINE=94117363; PubMed=8288523;

RA Akenbauer R.G., Quan H.N.;

RT "Fpta, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a

RT phenolate siderophore receptor homologous to hydroxamate siderophore

RT receptors.";

RL J. Bacteriol. 176:307-319 (1994).

RN [2]

RP SEQUENCE FROM N.A.

```

RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: HIGH-AFFINITY OUTER MEMBRANE RECEPTOR REQUIRED FOR THE
CC -!- TRANSPORT OF FE(III)-PYOCHELIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U03161; AAC43213.1; -.
CC EMBL; AB004839; AAC07609.1; -.
CC PIR; A36942; A36942.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rec; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 720 FE(III)-PYOCHELIN RECEPTOR.
FT SITE 703 720 TONB C-TERMINAL BOX.
FT SEQUENCE 720 AA; 79992 MW; DA796313116E0C2 CRC64;
Query Match 3.3%; Score 134.5; DB 1; Length 720;
Best Local Similarity 19.8%; Pred. No. 0.12;
Matches 174; Conservative 85; Mismatches 276; Indels 345; Gaps 43;
QY 1 MAQTTLKPIVLISILLINPLLAQAHTETQSVGLT-----VTVVGK-- 41
DB 1 IARNRHTFCLGLLALSLAFAAADAADKDGTELPLDMVISGESTSATQPPGVTTLGKVP 73
QY 42 ----SRPRATGSLHTSTASDKIISGDTLRQAVNL-----GDAL 77
DB 74 LKPRELPQASVIDHERLEQQNLFSLIDEAMQATGVTVPFOLLTTAYYVRGKVDSEFL 133
QY 78 DGVPGIHASOYGGGASAPVIRQGTGRIRKVLNHGCTGDMADFPDHALMVDYALSQQVE 137
DB 134 DGVPAL-----LGNATSSP-----QDMAIYERVE 157
QY 138 ILRGVPTLLYSNGVAGLVADVADGKIPKMPENGVSGBGLRLSSGNLEKLTSGGINIGL 197
DB 158 ILRGSNGLHGNGNPAATVNLV-----RKRQREFAA--STTLSAGRMDR----- 200
QY 198 GKNFVLHTEGLRGSGDVAVPRYRNKRLPDSRPFANGQRAVLGWRKRFYRTYSRR 257
DB 201 -----YRAEDVVGGLPSAS-----GNVRGRAVAAYEDRDY--FYDVA 235
QY 258 DQ-----YCLPAHSHEYDDCHADIWQSLINKRYLQL--YPHLLTREDV-----DY 302
DB 236 DQGTLLLVGT-----EFD-----LSPDTLLTVGAQVQHIDSTINMAGVPMKDG 280
QY 303 DNPGLSCGFHDDDDAHAAHNGKPIDLRNRYELRAEWKQPPFGFEALRVHLNRNDYHH 362
DB 281 SNLGLSRDYLVD-----ND-----REFW-DYRAFGLJ----- 309
QY 363 DEKAGDAVENFNNOTQVARIELRHQPIGRKLSGNGV--QYLGKSSALSATSEAVKQPM 420
DB 310 EQQLGGGWKGKVSABYQADSLRY-----AGSFGAIDPQTGGGQLMGAA----- 355

```

```

QY 421 LLDNKVQHYSPFGVQANWNTFLEGGYRVEKQKASIRYDKALIDRENYKQPLPLDGAH 480
DB 356 -----YKFKSIQORSLDAN--LNGPVL-----FGLTHELLGGVTYAQ-----GET 393
QY 481 RQ-TARSPALSG---NWYFTPOHKLSTLASHOERLPSTQELYAHGKHVATNTEVGNKHL 536
DB 394 QDTRAFNLNPTPNVNVYRWDPHGVPRPQIGQYTSFGT-----TTTOKGLYAL 442
QY 537 NK-BRSNNIELALGYEGDRQWQNLALYRNFGNYIYAQTLDNDRGPKSTEDSEMKLVRY 595
DB 443 GRIKLAEBLTLVVGRESWMDQDTPATRFKGR-----QFTPY 480
QY 596 NQSGADF-----YGAEGEYIFKPTPR-----YRIGVSGDYVRGLKNLP 634
DB 481 GGLIWFDFARDWSYVSYAEVTPQADROTWNSEPLSPVEGKTYETGIKGLADGRL-NLS 539
QY 635 SLPGREDAYGNRPFFIAQDDQNAVRP-----AARLGHFKAS--LTD--RIDANL 680
DB 540 LAAFRIDLNN-----PQEDPDHPGPPNPNPFVISGGKVRSQGFELEGTGYLTPYWSLSAGY 595
QY 681 DY---YRVFAQNKL-ARVETRTPGHMLNLGANY-----RR----- 712
DB 596 YTTSTEYLNKDSNDGSTRYSITFTP-RHLRLWSNTDLPWQDRRWSVGGGLQAQSDYSVDY 654
QY 713 -----NTRYG-----EWNWYVKADNLLNQSVY 734
DB 655 RGVSMRQGGYALVNNRLGKVIDEHWTAANVNNLFDRTVY 694

```

```

RESULT 28
PUPB_PSEPU STANDARD; PRT; 809 AA.
ID AC P38047;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ferric-pseudobactin BN7/BN8 receptor precursor.
GN PUPB.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RX MEDLINE=93316856; PubMed=8392140;
RA Koster M., van de Vossen J., Leong J., Weisbeek P.J.;
RT Identification and characterization of the pupB gene encoding an
RT inducible ferric-pseudobactin receptor of Pseudomonas putida
RT WCS358.";
RL Mol. Microbiol. 8:591-601(1993).
CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORES FERRIC
CC PSEUDOBACTIN BN8 AND FERRIC PSEUDOBACTIN BN7, IRON CHELATING
CC MOLECULES THAT ALLOW THE ORGANISM TO EXTRACT IRON FROM THE
CC ENVIRONMENT, SPECIALLY UNDER IRON-RESTRICTED CONDITIONS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY IRON LIMITATION IN ADDITION TO THE PRESENCE OF ONE
CC OF THE TWO COGNATE PSEUDOBACTINS BN8 OR BN7.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73598; CAA51995.1; -.
CC PIR; S32899; S32899.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rec; 1.

```

```
DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE_NEG.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.
FT SIGNAL 1 45 POTENTIAL.
FT CHAIN 46 809 FERRIC-PSEUDOBACTIN BN7/BN8 RECEPTOR.
FT DOMAIN 156 160 POLY-SER.
FT SITE 732 809 TONB C-TERMINAL BOX.
SQ SEQUENCE 809 AA; 88389 MW; 0B339F6E788A8C0D CRC64;

Query Match
Best Local Similarity 3.3%; Score 132; DB 1; Length 809;
Matches 133; Conservative 98; Mismatches 247; Indels 224; Gaps 34;

QY 11 LSLILINTPLLA-----QAHETEQSQVLETVTVVGVKSRPRATSG-LHTSTASDK 59
DB 100 LAILLAGTGLEASRGANASYSLOASASTGALSASVSGKAPGSTTEGTGLITYSSSS 159
QY 60 -----LISGDTL-RQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTG 102
DB 160 STRLNLTPRETQSLTWTRQRLDQRLNTLDALDATPGITVVRDGLGSESD---SYMS 216
QY 103 RRIKVLNHGGE---TGDMAFSPDHAIMVDYVDTALSOQVEILRGFVTLVSSGNVAGLVDA 159
DB 217 RFAIQNYEVDGVPSTRLDNYQSWMAFD-----RVELVRGATGLISGMGNFSAIINLI 271
QY 160 DGKIPEKMPENGVSGLGRLSSGNLEKLTSGGINIGLKNFVLHTEGLYRKSG----- 213
DB 272 -RKRPATAEAQAISITGE-----AGNWDRYGTG-----FDVSGPLTETGNIIRGRF 313
QY 214 --DYA-----VPRYRNKLKLP-----DS-----PRRAN 235
DB 314 VADYKTEKAWIDRYNQSQSLMGTEFDLSEDTLTVGFSYLRSDIDSPRLSGLTRFST 373
QY 236 GQH-----RAVLGKRKRYFART--YSRRDRQYGLPAHSHEYDDCHADIIWQKSLINKR 286
DB 374 GERTNLKRSINAAPDMSYNDHEQTSFPTSIEQQLG-NWMSGKIELTHAE-----NK 423
QY 287 YIQLYPHLITEEDVDYDNFGLS-----CGFHDDDDAHAAHNGKFWIDLRNKRVEL--- 337
DB 424 FDELFPFAMGE--LNPDPGSLQLPVRFGTFRQNDLIDYATGP--FSLFGREHLITG 478
QY 338 --RAEWKQPPGPEALRVHLNRNDYHDEKAGDAVENFNNTQNAIRLHQPIGRLLG 395
DB 479 MTLQSVQRENTPSWGGNR-----YDYAGSPAGIDNLFNWDGKSAK-----PAFVBSG 525
QY 396 SWGV---QYLGQKSSALSATSEAVKQPMLLDNKVQHSYFPGVEQAWNDFTLE---GGVR 449
DB 526 KSSIDEDQYAAYLTSRFSVTDDL---SLILGSL-----INWKEDTSDRPGGEE 572
QY 450 VEKQKASIRYDKALIDRENYKQPLDGAHRQTARSFALSGNWYFTPO-----498
DB 573 TEVNR-----EENGVFIPYAGVGVDLDDTWLSLYASYTKIPNPGAWVTDESINKP 621
QY 499 -----HKLSLTASH-----QERLPSTQ--ELYAHGKHVATNTFEVGN 533
DB 622 LDPMEGVGVELGKGTGLNKLNSLAVFKLEQDNLAIWQHDNVYSAEQDTTSKGIEL-- 679
QY 534 KHLINKERSNNIELALGY-----EGDRWQYNLALYENREGNY 569
DB 680 -ELNGELAEQWQASAGYSVTTDADDQRIINTLNLPNSFKTF 720

RESULT 29
HTPG VIBCH
ID_HTPG VIBCH STANDARD; PRT; 635 AA.
AC P23259; Q9KTB8;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein htpg (Heat shock protein htpg) (High temperature
DE protein G).
GN HTPG OR VC0985.
OS Vibrio cholerae.
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN 111
RC SEQUENCE FROM N.A.
RC STRAIN=El Tor M16961 / Serotype O1;
RC MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F.; Eisen J.A.; Nelson W.C.; Clayton R.A.; Gwin M.L.;
RA Dodson R.J.; Haft D.H.; Hickey E.K.; Peterson J.D.; Umayam L.A.;
RA Gill S.R.; Nelson K.E.; Read T.D.; Tettelin H.; Richardson D.;
RA Ermlaeva M.D.; Vamathevan J.; Bass S.; Olin H.; Dragoi I.; Sellers P.;
RA McDonald L.; Uterback T.; Fleischmann R.D.; Nierman W.C.; White O.;
RA Salzberg S.L.; Smith H.O.; Colwell R.R.; Mekalanos J.J.; Wenter J.C.;
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RT Nature 406:477-483(2000).
RL 21
RN SEQUENCE OF 1-110 FROM N.A.
RP STRAIN=El Tor E7946;
RC MEDLINE=91088618; PubMed=2124707;
RA Parsot C.; Mekalanos J.J.;
RT "Expression of ToxR, the transcriptional activator of the virulence
RT factors in Vibrio cholerae, is modulated by the heat shock
RT response.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:9898-9902(1990).
RL CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- INDUCTION: BY INCREASE IN TEMPERATURE
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AE004180; AAF94146.1; -.
CC EMBL; M58033; AAA27574.1; ALT_INIT.
CC PIR; B82255; B82255.
CC HSSP; P07900; 1YER.
CC TIGR; VC0985; -.
CC HAMAP; MF_00505; -.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR001404; Hsp90.
CC Pfam; PF02518; HATPase_C; 1.
CC Pfam; PF00183; HSP90; 2.
CC PRINTS; PR00775; HEATSHOCK90.
CC SMART; SM00387; HATPase_c; 1.
CC PROSITE; PS00298; HSP90; 1.
KW Chapterone; ATP-binding; Heat shock; Complete proteome.
FT DOMAIN 1 344 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 345 561 B (BY SIMILARITY).
FT DOMAIN 562 635 C.
SQ SEQUENCE 635 AA; 72197 MW; A635C6A39C151AAC CRC64;

Query Match
Best Local Similarity 3.2%; Score 131; DB 1; Length 635;
Matches 154; Conservative 95; Mismatches 282; Indels 286; Gaps 40;

QY 34 ETVTVGVKSR--PRATSGLLHT---STASDKIISGDTLRQKAVNLGDALDGVPGIHASQY 88
DB 3 ETATNTKTRGTFQSEVKQLLHMTHTSLYSNKEI---FLRELISNASDAVDKL----- 51
QY 89 GGGASAPVIRGQTGRRIKVLNHGEGTGMADFSPDHAIWDTALSOQVEILRGFVTLVSS 148
DB 52 -----RFAQLSH-----PD-LYQGDALGVKLSFDKQNTLTIS 84

149 SGNVAGLVADVADGKIPEKMPENGVSGLGRLSSGNLE---KLTS-----GGINIG 196
```

```

Db      85  DNGI-----GWTREVIEN-----LGTIAKSGTAEFKSLQSQKNSQILQFGVGG 131
QY      197  LGKNFVL-----HTEGLYRKSGDYAVPRYRNRLKRLPDSPPRFAN 235
Db      132  FYSAFIVADAVTVTRAAGSAPADAVQWYSGK-----EGEYVETINKESRGTDIILHLRE 187
QY      236  GQHRVAVLGHKPKFYRTYSOR-----RDQYGLPAHSHEYDDCH-ADIWOKS-- 281
Db      188  EGKEFLSEWRLRDVISKYSDHIGIPVYIQTSMDEGKATEETKWPQINKAQAALWTRAKS 247
QY      282  -LINKRYLOLYPHLLTEEDVDVNDPGLSCGFHDDDAHAHNN--GK----- 325
Db      248  EYDEYKFKYKLV-----SHDFADPLVWSHNVKVGKNDYTSLLYIPAK 291
QY      326  -PWIDLNKRYLAERAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNNTQONAR-- 382
Db      292  APW-DLFNREHK-----HGLKLYVQR-VFIMDDAA-----QFMPSYLRFVRGL 332
QY      383  IELRHQPIGELKSGWGVQVILGQKSSALSATSEAVKQPMLLDNKVQHYSPFGVEQANWDPF 442
Db      333  IDSNDLPLN-----VSREILODNKITQ-----SLRQA----- 359
QY      443  TLEGGVVRVEKQASIRYDKALIDRENKYK-----QLPLDLGAHQRTARSFALSG 491
Db      360  -----CTKRVLTMLERASNDADNYQKFWKEFGVLMKEGPAEDPANREKTASLLRPAS 412
QY      492  NWYFTQHLKSLTASHOERLPSTQE-----LYAHGKHVAVNTFVGNKHLNKNERNIELA 547
Db      413  THIDSAEQITSL-ASYVERMEKQDKIYVLTADSYTAANK-----SPHLEQFKSGKIEVI 466
QY      548  LGYEG-DRWOYNLALVRNFRGNYIYAQTLNDGRGPKSIEDSEMKLVRYNQSGADFYGAE 606
Db      467  LMFDRIDEWLMN--VLPEFGKAFOSITKAGLDLSQFDEAEKEKHK-----E 512
QY      607  GEIYFPTPRYRIGVSGDYVRG-----RLKNLPSLFGRED-AYGNR--PFIAQDDQNA 558
Db      513  TEEQFKSVVERLKGYSRVEKVRTFTFKLANTPAVVVTDYEMGTQWAKLLAAAGQVPE 572
QY      659  VPAARLGFHLKASLTDRIANDLYYVFAQNKRLARYETPTGHHMLNGLANVRNTRYGE 718
Db      573  VKYI-----LEVN-----PEHALVKRMADEAQTGFR 600
QY      719  WNWYKADNLNOSVVAHSSFLSDTPQMGRSFTGGYN 755
Db      601  W-----AEVLLGQMLAERGSMDPSQ-----FLGAVN 628

RESULT 30
PUPA_PSEPU
AC PUPA_PSEPU STANDARD; PRT; 819 AA.
AC P25184;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Ferric-pseudobactin 358 receptor precursor.
GN PUPA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RC MEDLINE=91260449; PubMed=1646376;
RX Bitter W., Marugg J.D., de Weger L.A., Tommassen J., Weisbeek P.J.;
RT "The ferric-pseudobactin receptor Pupa of Pseudomonas putida WCS358:
RT homology to TonB-dependent Escherichia coli receptors and specificity
RT of the protein."
RL Mol. Microbiol. 5:647-655(1991).
CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
CC PSEUDOBACTIN 358.
CC -!- SUBCELLULAR LOCATION: Outer membrane.

```

```

CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56605; CAA39942.1; --
DR PIR; S15169; S15169. TonB boxC.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB Dep Rec; 1.
DR PROSITE; PS00430; TONB DEPENDENT REC_1; 1.
DR PROSITE; PS01156; TONB DEPENDENT REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.
FT SIGNAL 1 47 POTENTIAL.
FT CHAIN 48 819 FERRIC-PSEUDOBACTIN 358 RECEPTOR.
FT SITE 115 122 TONB BOX.
FT SITE 802 819 TONB C-TERMINAL BOX.
SQ SEQUENCE 819 AA; 91015 MW; 1B36164F86207951 CRC64;

Query Match 3.2%; Score 130; DB 1; Length 819;
Best Local Similarity 20.7%; Pred. No. 0.3;
Matches 168; Conservative 101; Mismatches 291; Indels 250; Gaps 44;

QY 9 IVLSILLINTPILAAQAHETEQSGLTETVTVVGKSRPRATSGLLHTTASDKIIS---GD 64
Db 130 IELSATVNS---AGLGETTEGTGSYTRV-----TSTATVNNLSIRETPQ 172
QY 65 TLR-----QKAVNLGDALGVPGIHAHQYGGGASAPVIRGOTGRIRIKVNLHHG--- 112
Db 173 TITVTRQRMDDQHLGSMNEVLTTQTGITMSQDGERFNYSRGS- -INIYQFGVTT 229
QY 113 -----ETGDMADSPDHAIMVDYALTSQVLEILGPVTLTYSSGNVAGLVADAKIEKMP 168
Db 230 YQDNQTRNMP-----STLMDVGLYDRIEIVRGATGLMTGAGDPSAVVNI-RKRPTREF 282
QY 169 ENGVSGLG-----LRSSGNLEKLTSGGINIGLKNFVL-----HT-----EGLY-- 209
Db 283 KSHIQAGVGSWDYRAEADVSGPLTDDG--RVGRFFAAKQDNHTFMDWTQDRDVLVGV 340
QY 210 --RKSGDYAVPRY-----RNLKRLPDSPPRFANGQHRVAVLGHKPKFYRTYSDRR---D 258
Db 341 VEADVTDTTVARFGIDRQTYKYVNGAPGVPIIYVNGQ-----PTNFSRSTSSDAWGYD 393
QY 259 QY-----GLP-----AHSHE-----YDCHADII--WOKSLINKXYLQLYPHLLTEEDV 300
Db 394 DYTNTYTFELQQLAHQWQFKLAAAYMDVDRDSDFSYSYSTTTNRSYLEL----- 443
QY 301 DYDNPLSCGFHDDDDAHAAHNGKFWIDLNRKRYELRAEWKQFPFGFEALRVHLNRNDY 360
Db 444 -DGSTEISAGI-----VTAKQHQKGVDA TLQGP-FQLLGQTHELIVGVN---YLEVENK 492
QY 361 HHDEKAGDAVENPF--NNQTONARIELRHQPIGRLKSGWGVQY-LGQKSSALSATSEAVK 417
Db 493 HRGDSGPDVNIINFYDWNQTPKP-----GDDEIIFGIQVNIISNRQS----- 533
QY 418 QPMLLDNKVQHYSPFGVEQAN-WDNFTLGGVVRVEKQASIRYDKALIDRENKYKQLPDP 476
Db 534 -----GYFVASRFLNTDLDLILGARA-----SNYRFDYALWFRIGN--EPAPY 574
QY 477 LGAHQRTARSFALSGNWYFTPQHLKSLTASHOERLPSTQELVYAHGKHV--AINTFVGNK 534
Db 575 KMVERGVWTPYA--GIVY-----DLTNEQSVYASVTDIFKPNQNVDTIGK 617
QY 535 HLNKNRSNNIELALGVGD-----RWQYNLALVRNFRGNYIYAQTLNDGRGPKSIEDSEM 590
Db 618 PLDPFGVKNYE--LGMKGFLEGRNLNIALYVWVRDN--LAESTNE-----VVPDSGG 667
QY 591 KLVRYNQSGADFYGAEGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPREDAYGNRPFTA 650

```

Db 668 LIASRAVDAETKGVDELSEVLPGWVFTGYSHTR-----TEDADGKR----- 712

Qy 651 QDDQNAPRVPARLGFHLKASLTDRIDANLDYRVFAQNKLYARVTRTFGH-HMLNLGAN 709

Db 713 ----LTPQLP-----MDTFRFWNTYRL-----PGWEKLTLCGG 742

Qy 710 YRRNTR-----YGEWNYVYKADNLLNQSYVA 735

Db 743 VWNKSKTLNFAYNNSHVTQDDYFVTSLMA 772

RESULT 31

FTFBL STRMU STANDARD; PRT: 1476 AA.

AC P08987; O69381; O69384; O69387; O69390; O69396;

DT 01-NOV-1988 (Rel. 09, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Glucosyltransferase-I precursor (BC 2.4.1.5) (GTF-I) (Dextranucrase)

DE (Sucrose 6-glucosyltransferase).

GN GTFB OR SMU.1004.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GS-5;

RC MEDLINE=87308013; PubMed=3040685;

RA Shiroza T., Ueda S., Kuramitsu H.K.;

RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";

RL J. Bacteriol. 169:4263-4270(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,

RC MT4467 / Serotype E, and MT8148 / Serotype C;

RX MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=UAI59 / ATCC 700610 / Serotype C;

RX MEDLINE=22295063; PubMed=12397186;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,

RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental

RT pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT

CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE

CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE

CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS

CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-

CC fructose + {(1,6)-alpha-D-glucosyl}(N+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- DISEASE: DENTAL CARIES.

CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA

CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES

CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH

CC FORMS OF GLUCANS.

CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-

CC BINDING PROTEIN FROM S.MUTANS.

CC -!- SIMILARITY: Contains 10 cell wall binding repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M17361; AAA88588.1; -

DR EMBL; D88651; BAA26101.1; -

DR EMBL; D88654; BAA26105.1; -

DR EMBL; D88657; BAA26109.1; -

DR EMBL; D88660; BAA26113.1; -

DR EMBL; D89977; BAA26119.1; -

DR EMBL; AE014940; AAN58705.1; -

DR InterPro; IPR002479; CW binding.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF01473; CW_binding_1; 1.

DR Pfam; PF02324; Glyco_hydro_70; 1.

KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries;

KW Complete proteome.

FT SIGNAL 1 34

FT CHAIN 35 1476

FT DOMAIN 35 1051

FT DOMAIN 1097 1476

FT REPEAT 1097 1130

FT REPEAT 1161 1470

FT REPEAT 1161 1210

FT REPEAT 1225 1275

FT REPEAT 1290 1340

FT REPEAT 1355 1405

FT REPEAT 1420 1470

FT VARIANT 62 62

FT VARIANT 65 65

FT VARIANT 68 68

FT VARIANT 78 78

FT VARIANT 86 86

FT VARIANT 89 89

FT VARIANT 168 168

FT VARIANT 276 276

FT VARIANT 399 399

FT VARIANT 474 474

FT VARIANT 512 512

FT VARIANT 519 519

FT VARIANT 701 701

FT VARIANT 708 708

FT VARIANT 938 938

FT VARIANT 952 957

FT VARIANT 963 964

FT VARIANT 968 970

FT VARIANT 1086 1086

FT VARIANT 1158 1158

FT VARIANT 1163 1163

FT VARIANT 1168 1168

FT VARIANT 1182 1182

FT VARIANT 1234 1234

FT VARIANT 1263 1263

FT VARIANT 1263 1263

FT VARIANT 1264 1264

FT VARIANT 1272 1272

FT VARIANT 1329 1329

FT VARIANT 1394 1394

FT VARIANT 1402 1402

FT VARIANT 1459 1459

FT CONFLICT 570 570

FT CONFLICT 800 817

FT CONFLICT 1310 1310

SV -> NT (IN STRAINS GS-5, MT4239 AND MT4467).

ADS -> VDG (IN STRAINS GS-5, MT4239 AND MT4467).

A -> T (IN STRAIN MT4239).

S -> N (IN STRAIN MT4239).

H -> Y (IN STRAIN MT4251).

E -> K (IN STRAIN MT8148).

Y -> C (IN STRAIN MT8148).

A -> P (IN STRAIN MT4239).

R -> P (IN STRAIN MT8148).

R -> H (IN STRAINS GS-5 AND MT4467).

Y -> H (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).

S -> G (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).

H -> Y (IN STRAINS GS-5 AND MT4467).

Y -> H (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).

S -> G (IN STRAINS GS-5, MT4239, MT4467 AND MT8148).

Y -> H (IN STRAIN MT4467).

R -> A (IN REF. 1).

ADQVRAASTAPSTDGK -> LKMFALRLARPHQMA (IN REF. 1).

H -> L (IN REF. 1).

SQ SEQUENCE 1476 AA; 165846 MW; 9C6E09F731B4CBFC CRC64;

Query Match 3.2%; Score 129; DB 1; Length 1476;
Best Local Similarity 20.6%; Pred. No. 0.84;
Matches 144; Conservative 89; Mismatches 236; Indels 230; Gaps 44;

QY 87 QYGGGASAPVIRGQTRRIKVLN--HHGTGDMADSPDHAMV-----DTALSOQVILRG 141
DB 895 RYDLGTSKNGTGTADDLKAIALKASKGKVMADVPDQMYAFPEKVVVTRVDKFGK 954
QY 142 PV-----TLLY-----SSGN-----VAGLDVADVAKIPE-----KWPENGVSGELGLRL 180
DB 955 PVEGSIKSVLYVADSKSGKQQAQYGGAFLELOAKYPELFARQISTGTFMDSVSKI 1014
QY 181 SSGNLEKLTSGGINIGLGNFVHLTEGLYRKSGDYAVPRYRL-----KRLPDSPPRFANQ 237
DB 1015 KQWS-AKYFNGTINILRGAGYVLK-----DQATNTYFNISDNKEINFLPKTLNQD 1064
QY 238 HRAVLGWRKRFY-----RRTYSDRRDQVGLPAHSHEDDCHADIIWQSLINKRY 287
DB 1065 SQVGSYDGKGYVYSTGYQAKNTTIFISGDKW-----YFDNNGYMTGAQSGINGVNY 1118
QY 288 -----LQLYPHLLTEEDVDYDNPGLSCGFDHDDDAHAHAHNGKPMWIDLRNKRYE----- 336
DB 1119 YFLSNGQLRLDALKNEDCTY-----AYGND-----GRYENGYIQ 1155
QY 337 -LRAEWKQFPFGEALRVHLNRND-----YHDE-----KAGDAVENF 373
DB 1156 FMSGVWR-----HFNNGEMSVGLTVIDGVQYFDEMGYQAKGFVTTADGKIRY 1204
QY 374 FNNQTCNARIELRHQIPGLRKSGWQVYLGKSSALSATSBAVKQPMPLDNKVQHSFPG 433
DB 1205 FPKQSGNM---YRNRFIENEKRW---LYLGEDAAVVG-SQTNG-----QHLYF--- 1248
QY 434 VEQANWDFLREGVVRVEKASIRYDKALIDRENTYKQFLPDLCGAHROTARSFA--LSG 491
DB 1249 --RAN-----GVQVKGEFVTDY-----GRISY-----DSNSGDQIRNRFVNAQ 1288
QY 492 NNYFPOHKLSTASHOERLPSTQELYAKGHVATNFTFVGNKHLNKRNSNIELALGYE 551
DB 1289 QWFYFDNNGYAVTGA---RTINGQHLYFRANGVQVKGEFVTDH-----GRISYDNGS 1339
QY 552 GDRWOYNLALYRNF-----GNVIYAQTLDNGRGPKSIEDDSEMKLVRYNQSGADPYGAE 606
DB 1340 GDQ-----INRFRVNAQGWYFED--NNG-----YAVTGARTINGQ 1374
QY 607 GBYFKPTPRYRIGSGDYVR---GRKLNPLPGREDAYGNRPFT--AQD-----DQNA 656
DB 1375 -HLXYFRAN---GVQVKGEFVTDYGRISYDNSG--DQIRNR-FVRNAQGWYFEDNNG 1427
QY 657 PRVPAARL--GPHL--KAS-----LTDRI DANLDYY 683
DB 1428 YAVTGARTINGQHLYFRANGVQVKGEFVTDY--GRISY 1465

RESULT 32
LBPA NEIMA
ID LBPA NEIMA STANDARD; PRT; 944 AA.
AC Q9JTK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin binding protein A precursor.
GN LBPA OR NMA1739.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022556; PubMed=10761919;
R1 Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491";
RL Nature 404:502-506(2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL162757; CAB84967.1; -.
DR PIR; C81798; C81798.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB_dep_rec; 1.
DR PROSITE; PS00430; TONB DEPENDENT_REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 27 POTENTIAL
FT CHAIN 28 944 LACTOFERRIN BINDING PROTEIN A.
FT SITE 827 944 TONB C-TERMINAL BOX.
SQ SEQUENCE 944 AA; 105732 MW; CE06B6192E74AE3E CRC64;

Query Match 3.2%; Score 127.5; DB 1; Length 944;
Best Local Similarity 19.7%; Pred. No. 0.56;
Matches 197; Conservative 117; Mismatches 335; Indels 351; Gaps 54;

QY 24 AHETQSQVLETVTV-----VGKSRPRATSGLLHTSTADKII-SGDTL-RQKAVNLGDAL 77
DB 31 ASDAAQSQSLKEITVRAAKVGR-RSKEATGL-----GKIVKTSLTNKEQVLGIRDLT 82
QY 78 DGVPGIHASOYCGGASAPV-IRGQTRRIKV-----LNHSETGMADP 120
DB 83 RYDPGVAVVEQNGASGGYSIRGVDKNRVAVSDVGAQIAFTVQGSLSGYGGRGGSGAI 142
QY 121 SPDHAIWVDLTALSQQVEILRGPVTLTYSSGNVAGLVD-----VADGK---IFEKMP 168
DB 143 NEIEYENIST-----VEIDKGAGSSDHGSGALGGAVAFRTKEAADLISDKSWGIQAKTA 197
QY 169 ENGVSQE---LGLRLSSGNLEKL-----TSGGINIGH----- 197
DB 198 YGSKNRQFMKSLGAGFSKDGWEGLLIRTERQGRETRPHGDIADGVEYIDRLDAFRQTYD 257
QY 198 -----GKNFVLHTE-----GLY-----RKSGDY----- 215
DB 258 IQQNKKAIFYLAEGEREKPKVAKLAGNGYILKNQLRNWRVBERKKNQPLNAEEAMVRE 317
QY 216 AVPRYRNL-----KELPDSPPRFANQGHRAVLGWR---KRFYRTYSRRRDQYGLP 263
DB 318 AQARHENLSAQSYVTGGRIIPD-PMDYRSGSWLAKLYRFGGRHYGVGFEDTKQRYDI- 375
QY 264 AHSHEYDDCHADIIWQSLINKRYQLYPHLLTEEDVDYDNPGLSCGFDHDDDAHAHAHN 323
DB 376 -----RDMTEKQYVG-----TDEAKKFSN---KSGVVDGN----- 402
QY 324 GKPMWIDLRNKRYELR--NEWKQFPFGFALRVHLNRN---DYHH----- 362
DB 403 -----DFRDGLYFVPEIIEWKGTNLVKIGLKYGRTRKFIDEHRRRRMGLLYRYENEAY 457
QY 363 -DEKAGDAVENFPFN--QTONARIEL-----RH 387
DB 458 SDNWADKAVLFDKGVATDNTNLKNCVAVPAVDKSCASADKFPYSVDSRDFHYRSGH 517


```

QY 536 LNKERSNNIELALGVGDRWQVNLALYRNRFGNYIYAOTLNDGRGPKSIEDDSEMKLVRY 595
Db 473 -----ELVVGTSAS-----FSHW-----EGKSYWNLRY 496
QY 596 NOSGADFYGAEGEIVFKPTPRYRIGVSGDYVRGRILNPLSLPGREDAYGNRPPIAQDDQN 655
Db 497 DNTTDDFINWGDIL-GKPD-----WGTPSQYIDDKTRQLGS-----YWTARFNVTTDD-- 542
QY 656 APRVPAARGLHKLASLTDRIANLDYRVFAQNKLARYETTPGHMLNLGANYRNRTR 715
Db 543 -----LNLFLGGRVVD-----YVETGLNP-----TIRESGRFPIYGVAVVDLNDT 582
QY 716 YGEW-----NWVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 583 YSVYASYTDIFMPQDSWYRDSNKLLE-----PDEGQYIEIGKGEY 624

```

RESULT 34

```

HTPG_SALTY
ID HTPG_SALTY STANDARD; PRT; 624 AA.
AC P59480;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein htpG (Heat shock protein htpG) (High temperature
DE protein G).
GN HTPG OR STM0487.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RA MEDLINE=21534948; PubMed=11677609;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Letreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008718; AAL19441.1; ALT_INIT.
CC StyGene; SG????; htpG.
CC HAMAP; MF_00505; -; 1.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR001404; Hsp90.
CC Pfam; PF02518; HATPase_c; 1.
CC Pfam; PF00183; HSP90_2.
CC PRINTS; PR00775; HEATSHOCK90.
CC SMART; SM00387; HATPase_c; 1.
CC PROSITE; PS00298; HSP90; 1.
CC Chapterone; ATP-binding; Heat shock; Complete proteome.
KW DOMAIN 1 336 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 337 552 B (BY SIMILARITY).
FT DOMAIN 553 624 C.
SQ SEQUENCE 624 AA; 71486 MW; 0974B2B1B38D0B39 CRC64;

```

```

Query Match 3.1%; Score 123.5; DB 1; Length 624;
Best Local Similarity 20.1%; Pred. No. 0.58;
Matches 150; Conservative 85; Mismatches 293; Indels 219; Gaps 32;

QY 54 STASDKIISGDTLRKAVNLGDAL--DGVPGIHASQYGGGASAPVIRGTG-RRIKVLNH 110
Db 37 SNASD---AALKLFRALSNFDLYBGGELRVRSFDDKRTLTITADNGVGMNRDEVIDH 93
QY 111 HGETGDMADFDPDHAIMVDTALSQQVEILRGFVTL-LYSSGNVAGLVADVADGKIPKQPE 169
Db 94 LGTAKSTKTFLESMSGDQAKDSQ---LIQFGVGFSAFIVADKVVTRRAADGK-PE 149
QY 170 NGVSGELGRLSSGNLEKLTSGGINIGLGNKFLVHLTEGLYKSGDYAVPRYRNLKRLPDS 229
Db 150 NGVFE-----SAG-----EGEYTVADITKDRGTETI 176
QY 230 PRRFANGQHRAVLGHRKFPYRTYSDDRDQVGLPA--HSHEYDDCHADIIWQKSLINK-- 285
Db 177 TLHLRGEDEFLDDMRVRSIIISKYS---DHIALFVEIEKREEKDGGETVISWEK--INKAQ 231
QY 286 -----RYLOLYPHLLTEEDVDYDNPGLSCGFHDDDDDAHAHAHNGKPWIDLRN 332
Db 232 ALWTNKKSEIKDDEYNEFYKHIA-----HDTDTPLTWSHN----- 266
QY 333 KRYELRAEWKQFPFGFEALRVHLNRNDYHHDEKAG-----DAVENPFNNQTQNA 381
Db 267 -----RVEGKQETSLLYIPSAQPDWDLNRDHGHLKLYVQRVFTMDDAEQMPNYLRFV 321
QY 382 R--TELHQPIG-----RLKSGWGVQYLCQKSALSATSEAVKQPMLLDNKVQHYHFFGVE 435
Db 322 RGLIDSNLPLNVSREILQDSTVTRNL---RSALTQRYVLMLEKLAQKADAKEYQTF---- 374
QY 436 QANWDNFTLEGVRVEKQKASIRYDKALIDRENYKQPLDPLGAHQRTARSFALSGNWFY 495
Db 375 ---WKQF---GLVLKEGPAEDHANQEAIALKLLRFASHTHTSSAQTSLEDY----- 419
QY 496 TPQHKLSUTASHQERL-PSTOELYAHGKHVATNTFEVGNKHLNKNERSNNIE-LALGYEGD 553
Db 420 -----VSRMEGQEKIYIITADSYAAAKN-----SPHLELRKKKGIEVLLLSDRID 465
QY 554 RWQYNLALYRNRFGNYIYAOTLNDGRGPKSI--EDDSEMKLVRYNQSGADFVGARGEIYF 611
Db 466 EWMVN---YLTEF-----DGKAFQSAKADESIEKLA-----DEVENAKSAEKAL 508
QY 612 KPTPRYRIGVSGDYVRG-----RLKNLPSLPGREDAYGNRPPIAQDDQNAPRVPAARLGF 666
Db 509 EPFVERVKTLTGDRVKDVLRLTLRLTDTAI----- 538
QY 667 HLKASLTDRIANLDYRVFAQNKLARYETR-----TPGHMLNLGANYRNRTRYGENWN 721
Db 539 -----VTTDAEMSTQMAKLFAAAGQSVPEVKYIFELNPDHVLVVKRTADTKDAQPKW-- 592
QY 722 YVKADNLLNQSVYAHSSFLSDTPQMG 748
Db 593 ---VELLDDQALFAERGTLDPNQFIR 616

RESULT 35
HTPG_SALTY
ID HTPG_SALTY STANDARD; PRT; 624 AA.
AC P59479;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chapterone protein htpG (Heat shock protein htpG) (High temperature
DE protein G).
GN HTPG OR SY0531 OR T2373.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.

```

Db	232	ALWTRNKS	I	K	D	E	V	E	F	F	K	H	A	---	H	D	F	D	P	L	T	W	S	H	---	266		
Qy	333	KRYELRAEWKQPPGPEALRVLHNRNDYHHDEKAG	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	381		
Db	267	----	R	V	E	G	K	Q	E	V	T	S	L	I	P	S	Q	A	P	M	D	L	N	R	D	---	321	
Qy	382	R	---	I	E	L	R	H	O	P	I	G	---	---	---	---	---	---	---	---	---	---	---	---	---	435		
Db	322	RGLISDNDLPLNVSRILOQDSTVTRNL	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	374		
Qy	436	QANWNTLEGGVVRVKQKASIRYDKALIDRENYKQPLDLCGHRQTARSFALSGNWF	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	495		
Db	375	----	W	K	Q	F	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	419		
Qy	496	TPQKLSLTASHOERL	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	553		
Db	420	----	V	S	R	M	K	E	G	O	E	K	I	Y	I	T	A	D	S	Y	A	A	K	N	---	---	465	
Qy	554	RMQYNLALVNRFRGNYIYAQTLDNGRGP	K	S	I	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	611		
Db	466	EMMN	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	508		
Qy	612	KPTFRYRIGVSGDYR	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	666		
Db	509	EPFVERKTLGDKRVEVRLTHRLTDP	A	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	538		
Qy	667	HLKASLTDRIDANLDYRVRFAQNKLARYETR	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	721		
Db	539	----	V	T	T	D	A	D	E	N	S	T	O	M	A	K	L	F	A	A	G	S	V	E	P	E	---	592
Qy	722	YVKADNLLNQSVYAHSSFLSDTPQ	M	G	R	748	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Db	593	----	V	E	L	L	D	Q	A	L	F	A	E	R	G	T	L	E	D	P	N	O	F	I	R	616		

RESULT 36

HTPG VIBPA

AC

Q87RH5;

ID

15-SEP-2003 (Rel. 42, Created)

DT

15-SEP-2003 (Rel. 42, Last sequence update)

DT

15-SEP-2003 (Rel. 42, Last annotation update)

DE

Chaperone protein htpg (Heat shock protein htpg) (High temperature protein G)

DE

HTPG OR VP0821.

OS

Vibrio parahaemolyticus.

OC

Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC

Vibrionaceae; Vibrio.

OK

NCBI_Taxid=670;

[1]

SEQUENCE FROM N.A.

RP

STRAIN=RMD 2210633 / Serotype O3:K6;

RX

MEDLINE=22508454; PubMed=12620739;

RA

Makino K., Oshima K., Kurokawa K., Yokoyama K., Uga T., Tagomori K.,

RA

Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,

RA

Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;

RT

"Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae.";

RL

Lancet 361:743-749(2003).

CC

!- FUNCTION: Molecular chaperone. Has ATPase activity (By similarity)

CC

!- SUBUNIT: Homodimer (By similarity).

CC

!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC

!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.isb-sib.ch/announcement>

CC

or send an email to license@isb-sib.ch.

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
[1]
NCBI_TaxID=1417;
SEQUENCE FROM N.A., AND SEQUENCE OF 28-47.
MEDLINE=90147765; PubMed=1689153;
Tukor P. Tsukagoshi N. Uda S.;
"Nucleotide sequence of the raw-starch-digesting amylase gene from
Bacillus sp. B1018 and its strong homology to the cyclodextrin
glucoamylase genes";
Biochem. Biophys. Res. Commun. 166:630-636(1990).
-!- FUNCTION: THIS ENDO-TYPE ADSORBABLE AMYLASE IS CAPABLE TO
DIGEST RAW-STARCH.
-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
-!- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
-!- SUBCELLULAR LOCATION: Secreted.
-!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
KNOWN AS THE ALPHA-AMYLASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M33302; AAA22239.1; -;
EMBL; D90112; BAA14140.1; -;
PIR; S09196; S09196.
HSP; P43379; 1CDG.
InterPro; IPR006589; Alp_aml_cat_sub.
InterPro; IPR006048; Alpha_aml_C.
InterPro; IPR006047; Alpha_aml_cat.
InterPro; IPR002044; CBD 4.
InterPro; IPR006046; Glyco_hydro_13.
InterPro; IPR002909; IPT_TIG.
Pfam; PF00128; alpha-amylase; 1.
Pfam; PF02806; alpha-amylase-C; 1.
Pfam; PF00886; CBM_20; 1.
Pfam; PF01833; TIG; 1.
PRINTS; PR00110; ALPHAMYLASE.
ProDom; PD001568; CBD 4; 1.
SMART; SM00642; Amy_1; 1.
SMART; SM00632; Amy_C; 1.
Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium; Signal.
FT SIGNAL 1 27
FT CHAIN 28 713 RAW-STARCH-DIGESTING AMYLASE.
FT ACT_SITE 256 256 BY SIMILARITY.
FT ACT_SITE 260 260 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
SQ SEQUENCE 713 AA; 77420 MW; 85FB616DA687B888 CRC64;

Query Match 3.0%; Score 120.5; DB 1; Length 713;
Best Local Similarity 18.5%; Pred. No. 1.1;
Matches 143; Conservative 105; Mismatches 273; Indels 253; Gaps 38;

QY 120 ESPDHAIWDTALSSQ-----VEILRGVPTLLSYSCNVAGLVADVADGKIPEKMPENGVSIGE 175
Db 22 FSPAQA-APDTSVNKNQNFSTDIYQIFTFDFSDGN-----PANNPFGA 64

QY 176 -----LGRLSSGN-----LEKLTSGLINIGLKFNFLHTTEGLYKSGDYAVPYRNLIK 224
Db 65 AFDGTCTNLRLYCGDWQGIINKINDGYLT-GMGVTAINISQVF---ENIYSINIISGV- 119

QY 225 RLDPSPRRFANGQHRAVLGWKRKYRET---YSDRRDQYGLPAHSHEYDDCHADIHWKS 281
Db 120 -----NNNTAYHYGWARDFKKTNPAYGTIADFQNLIATAAH-----AKN 156

QY 282 LINKRYQLQYPLHLTEEDVD-----YDNPGLSGCFHDDDDHAHAHNKPKWIDL 331
Db 157 I--KVIFDAPNHTSPASSQQPSFAENGRLYDNGTLGGYTNQTOLFHNGTGDFSTTE 214

CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1417;
RN SEQUENCE FROM N.A., AND SEQUENCE OF 28-47.
RP MEDLINE=90147765; PubMed=1689153;
RA Tukor P. Tsukagoshi N. Uda S.;
RT "Nucleotide sequence of the raw-starch-digesting amylase gene from
Bacillus sp. B1018 and its strong homology to the cyclodextrin
glucoamylase genes";
RL Biochem. Biophys. Res. Commun. 166:630-636(1990).
CC -!- FUNCTION: THIS ENDO-TYPE ADSORBABLE AMYLASE IS CAPABLE TO
CC DIGEST RAW-STARCH.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M33302; AAA22239.1; -;
EMBL; D90112; BAA14140.1; -;
PIR; S09196; S09196.
HSP; P43379; 1CDG.
InterPro; IPR006589; Alp_aml_cat_sub.
InterPro; IPR006048; Alpha_aml_C.
InterPro; IPR006047; Alpha_aml_cat.
InterPro; IPR002044; CBD 4.
InterPro; IPR006046; Glyco_hydro_13.
InterPro; IPR002909; IPT_TIG.
Pfam; PF00128; alpha-amylase; 1.
Pfam; PF02806; alpha-amylase-C; 1.
Pfam; PF00886; CBM_20; 1.
Pfam; PF01833; TIG; 1.
PRINTS; PR00110; ALPHAMYLASE.
ProDom; PD001568; CBD 4; 1.
SMART; SM00642; Amy_1; 1.
SMART; SM00632; Amy_C; 1.
Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium; Signal.
FT SIGNAL 1 27
FT CHAIN 28 713 RAW-STARCH-DIGESTING AMYLASE.
FT ACT_SITE 256 256 BY SIMILARITY.
FT ACT_SITE 260 260 BY SIMILARITY.
FT ACT_SITE 355 355 BY SIMILARITY.
SQ SEQUENCE 713 AA; 77420 MW; 85FB616DA687B888 CRC64;

Query Match 3.0%; Score 122; DB 1; Length 634;
Best Local Similarity 17.9%; Pred. No. 0.75;
Matches 128; Conservative 78; Mismatches 231; Indels 280; Gaps 34;

QY 153 AGLVADVADGKIPEKMPENGVSSELGLRLSSGNLE---KLTS-----GGINIGLGN 200
Db 78 ANTLTISDNGI--GMSRNDVIHELGTIAKGPAAFFSKLSEEQSDSLQIQGFVGFSYA 135

QY 201 FVL-----HTEGLYKSGDYAVPYRNLR-----LPDSPPR 232
Db 136 FIVADAVTVTRAAGLPADAENVQHSAG---EGEYTIENITKESRGTDIIHLWRDGEKE 191

QY 233 FANGQRAVLGWKRKYRTYQSR-----RDQYGPAHSHEVDDCH-ADIHWK 280
Db 192 FLN-----EWRLRDVISKYSHDGIPVSIOTVVVRDEDKETDEKKWEQINKAQALWTR 244

QY 281 S---LINKRYQLQYPLHLTEEDVDYDNPGLSCGFHDDDDHAHAHNKPKWIDLNKRYEL 337
Db 245 NKADISDEBYQFYKHV-----SHDFADPLVWSHR--VEGKNDVTSL 285

QY 338 ----RAEKQPFPGFEALRVHLNRNVDYHDEKAG-----DAVENFFNNQTONA 381
Db 286 LVPISKAPDM-----MNR-----DHKSLGLYQVRFIMDDAEQFMPSYLRFV 329

QY 382 R---IELRHQPIGELKSGWGVOYLQKSSALSATSEAVKQPMLDNKV-----426
Db 330 RGLIDSNDLPLN-----VSREILODNKVYQSRLNACTXRVL 365

QY 427 ----OYISFFGVEQANWDNFTL-----EGGVREVKKASIRYDKALIDREN 468
Db 366 TMLERMAKNDKEYQSF-----WKFEGLVLKEGPAEDFANKETIAGLLRFASTEVD-- 416

QY 469 YKQPLPDGAHQROTARSFALSGNWYFPPOHKLSITASHQRRLPSTQE----LYAHGKHV 524
Db 417 ----SAEQTVG-----LASVVERMKEGQDKIYVLTAASYAA 448

QY 525 ATNTPEVGNKHLNKRNSNIIEALGVG-DRWOYNALYRNFRNGNYIAQTINDRCRPSK 583
Db 449 AKN-----SPHLEQPKAKGIEVILPDFRIDEWLN---YLTEFGKGQFQSTIKAGLDLSK 500

QY 584 IEDDSMKLVRYNQSGADPYGAGGIYFKPTPRVRIAGSVGDIVRG-----RLKNLPLSPG 638
Db 501 FEDEADKERQK-----ETEDEFKSVVERTKSYLGDVRVKDVRTTKLASTPAV-- 547

QY 639 REDAYGNRFFIAQDONAPRPAARLGHPLKASLTDRIDANDLYRYPVPAQKLARYETR 698
Db 548 -----VTTDY-----EMGTOMAKLIAAQAQVPEVKYIFEIN-----580

QY 699 PGHHMLNLGANRYRTRYGENWYVYKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVN 755
Db 581 PEHELVKRMADAEEAFGRW-----VEVLIGQAMLAEBSGWEDPTQ-----FLGAIN 628

RESULT 37
AMR_BACS8
ID AMR_BACS8 STANDARD; PART; 713 AA.
AC PI7692;
DT 01-AUG-1990 (rel. 15, Created)
DT 01-AUG-1990 (rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, last annotation update)
DE Raw-starch-digesting amylase precursor (EC 3.2.1.1) (1,4-alpha-D-
glucan glucoamylase).
OS Bacillus sp. (strain B1018).

```

QY 332 N----KRYELRAEWKQPPGPEALRVHLNRNDYHDEKAGDAVENFFNNQTONARIE-LR 386
Db 215 NGIYKNLHDL-----ADLNHNSSTDVYLKDKAKWLDLIGDIGIRMDAVK 259
QY 387 HOPICRLK-----GSW--GVQYLGQKSSALSATSBAVKQPMLLDNKVRQY 429
Db 260 HMPFGWQKSFMAAVNNYKPVFTFGWFLGVNVEGPNENHKFANESGMSLLDPRFAKVQ- 318
QY 430 SFFGVEQANWDF-----TLEGG-----VRVEKQKASI-RYDKALIDRENNYKQIPDLGA 479
Db 319 ----VFRDNTDMYGLKAMLEGAADYQVDDQVTFIDNHDMERFHASNANRKL----- 369
QY 480 HROTARSPAL-----SGNNYFTPOHKLST-ASHOERLPSTQELVYAHGKHVATNTFEVGN 533
Db 370 --EQLAFLLILARPAIYGTQYMSGGTDPDNARIPFS-----TSTTAYQVIO 419
QY 534 KHLNKRNNIBALGYEGDRQYNN-LALYRNRFGNYIYAQTLNDG-RGPKSIED-DSEM 590
Db 420 K-LAPLRKSNPAIAYGSTQERINNDVLIYERKFGSNVAVVNRNLNAPASISGLVTS 478
QY 591 KLVRYNQ----- 597
Db 479 POGSYNDVLGGLLNGTITVSGGGAASFTLAAGTAVWQYTAATPTIGHVGPMAKP 538
QY 598 ----SGADFYGARGEIYFPTPRYRIGVSGDYVRGRKLNPLSLPGREDAYGNRPFFIAQ 651
Db 539 GWTIITIDRGFGSSKGTYYFGTT-----AVSG-----ANITSW-----E 572
QY 652 DDQNAIPVPAARLG-FHLKASLUTRIDANL-DYRVVFAONKLA-----RYEFTPTGHEML 704
Db 573 DTQIKVKIPAVAGGIYNIKVANAAGTASNVDYFVLSQGVSRVFNATLQGLY 632
QY 705 NLGANVRENTRYGWNVYVKADNLNQSVAHSSFLSTPQMGSRFTGQVNVKF 758
Db 633 LTG-----NVSELGNWDPAKAGPMNQVYQIPNMYDV-----SVPAGKTIEF 677

```

RESULT 38

```

ID YDDB ECOLI STANDARD; PRT; 790 AA.
AC P31827; P77572;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yddB (CDS103).
GN YDDB OR B1495.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12.
RA Tulin E., Gasser F., Biville F.;
RT "Sequence and functional analysis of an Escherichia coli DNA fragment
RT able to complement pqeE and pqgF from Methylobacterium organophilum.";
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,

```

```

RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasundaram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377 (1996).
CC -!- SIMILARITY: TO H.INFLUENZAE HT1369.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL, X71917; CAA50733.1; -.
DR EMBL; AE000246; AAC74568.1; -.
DR EMBL; D90791; BAA15166.1; -.
DR PIR; B64903; B64903.
DR EcoGene; EG11743; YddB.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rc; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 1 22 MKRVILPGVILGADVAQAVDD -> MT (IN REF. 1).
FT CONFLICT 279 289 TWFSADRTHD -> YLVCORPLYPR (IN REF. 1).
FT CONFLICT 503 503 A -> T (IN REF. 1).
FT CONFLICT 534 545 TLTRYQDLKTPY -> NSDALSGFE (IN REF. 1).
FT CONFLICT 573 574 HD -> LV (IN REF. 1).
FT CONFLICT 615 616 HI -> LY (IN REF. 1).
SQ SEQUENCE 790 AA; 89282 MW; 907735A2A124E674 CRC64;

Query Match          3.0%; Score 120.5; DB 1; Length 790;
Best Local Similarity 19.8%; Pred No. 1.3;
Matches 175; Conservative 104; Mismatches 333; Indels 271; Gaps 47;

QY 16 INTPLLAQAHTEQSVGLTETVTYVVGKSRPRAT---SGLLHTSTASDKIISGTLRQAVN 72
Db 39 VVPVNGNTHYTSESI-----ERLPTGNGNISLLRTNPA---VRMDSTQSTSLN 85
QY 73 LGDALDGVPGIH-ASQYGGGASAPVIRGTGRRIKVLNHHGETGMADPSP-----DH 124
Db 86 QGDIRPEKISIHGASPY---QNAVILDG-----ISATNNLNPNAN-SDASSATNIGMSQ 136
QY 125 AIMVDYALSQQVEILRGVPTLYLSSGNVAGLVDA---DGKI-----PE 165
Db 137 GYLLDVSLDNTLYDSFVPEVGFEN-GVIDAKIKFNADDKVKLGRTTRSDWLTS 195
QY 166 KMPENGVS---GELGLSLSSGNLEK-LTSGGINIGLGNFVLHTEGLYRKSGDYAVPRY 220
Db 196 HIDDENKSAFNQSGSGSTYSPDFKKNFYTLISFNQELADNFGV-TAGLSRRQSDITRADY 254
QY 221 -----RNLKELPD-----SPRFANGQHRVILGWRKFRVRY--SDR--- 256
Db 255 VSDGIVAGRAQYKNVIDTALSFTWFASDRFTHDLTLKYTGSSRDYNTSTFPQSDREMG 314
QY 257 RDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTBEDVDYDNPGLSCGFHDDDD 316
Db 315 NKSYGLAW-----DMDTQLAWAKLRTTVGVHDHSDIYTRHDHDIWYTE--LSCITYGDI 366
QY 317 AHAHANGKPIDLRNRYELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFFNN 376
Db 367 RCTRGGLGHISQAVDNTVTFKRLDW-QKF-----AVGN----- 398
QY 377 QTONARIELRHOP-IGR---LKGSGVQYLGQKSSALSATSBAVKQPMLLDNK----VQH 428
Db 399 -----VSHQPYFGAEIYSDAWTERHNSQESVYVNAAGKTNHTIYHKGKRLGIDN 450
QY 429 YSFFGVEQANWDFTELEGVREKQKASIRYDKALIDRENNYKQIPDLGHRHQTARSFA 488
Db 451 YTLVMDRISWRNVSLMPGV-----YDYN-----LSNHNISPRFMT 489

```


Result No.	Query No.	Score	Query %			DB	ID	Description
			Match	Length				
1	3845	95.3	764	16	Q9JUS2	Q9JUS2	neisseria m	
2	3836	95.0	758	16	Q9JZN9	Q9JZN9	neisseria m	
3	1724	42.7	765	2	Q9ASR9	Q9ACR9	pasteurella	
4	1492.5	37.0	809	16	Q9CLW8	Q9CJW8	pasteurella	
5	1364	33.8	805	16	Q9CLZ0	Q9CJZ0	pasteurella	
6	1307	32.4	925	16	Q9CMR7	Q9CJL20	pasteurella	
7	745	18.5	688	16	Q8P7F3	Q8P7F3	xanthomonas	
8	718.5	17.8	687	16	Q8PIR6	Q8P7F3	xanthomonas	
9	694.5	17.2	681	16	Q9PCB8	Q8PIR6	xanthomonas	
10	679	16.8	687	16	Q9P5F7	Q9PCB8	xylella fas	
11	636	15.8	687	16	Q9ABL3	Q9P5F7	pseudomonas	
12	395.5	9.8	777	16	Q8F1I6	Q9ABL3	caulobacter	
13	354.5	8.8	778	16	Q9CLF3	Q8F1I6	leptospira	
14	290	7.2	713	16	Q9KNA1	Q9CLF3	pasteurella	
15	264.5	6.6	706	2	Q93TP3	Q9KNA1	vibrio chol	
16	264.5	6.6	715	16	Q8FGA0	Q93TP3	escherichia	
							Q8FGA0	escherichia

Db 7 MAQTTLKPIVLSILLINTPLLSQAHTQEQSGLETVSVVGKSRPRATSGLLHTSTASDKI 66
QY 61 ISGDTLRKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADF 120
Db 67 ISGDTLRKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADF 126
QY 121 SPDHAIMVDTALSQQVEILRGVPTLLYSYSGNVAGLVDVADGKIPEKMPENGVSGLGLRL 180
Db 127 SPDHAIMVDSALSQQVEILRGVPTLLYSYSGNVAGLVDVADGKIPEKMPENGVSGLGLRL 186
QY 181 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVPRYRNKRLPSPRFANGQHRA 240
Db 187 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVPRYRNKRLPSPRFANGQHRA 243
QY 241 V-LGW--RKRFRYRYSRRDOYGLPAHSHVEDDCHADIWOKSLINKYLOLYPHLLTE 297
Db 244 IGLSWVGKGFAGYSDRRDQYGLPAHSHVEDDCHADIWOKSLINKYLOLYPHLLTE 303
QY 298 EDVDYDNPGLSCGFFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 357
Db 304 EDIDYDNPGLSCGFFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 363
QY 358 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRKSGWGQYLGQKSSALSATSEAVK 417
Db 364 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRKSGWGQYLGQKSSALSATSEAVK 423
QY 418 QPMLLDNKVQHSYFFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYKOPLPDL 477
Db 424 QPMLLDNKVQHSYFFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYKOPLPDL 483
QY 478 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
Db 484 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 543
QY 538 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQ 597
Db 544 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQ 603
QY 598 SGADFYGAEGEYIFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFFAODDQAP 657
Db 604 SGADFYGAEGEYIFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFFAODDQAP 663
QY 658 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANYRNTRYG 717
Db 664 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANYRNTRYG 723
QY 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 724 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 764

RESULT 2

Q9JZN9 PRELIMINARY; PRT; 758 AA.
ID Q9JZN9
AC Q9JZN9
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TonB-dependent receptor.
GN NMB0964
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,

RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiangani V., Pizzo G., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
MC58.";
RL Science 287:1809-1815 (2000).
DR EMBL; AB024448; AAF62323.1; -.
DR TIGR; NMB0964; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 758 AA; 85119 MW; 6BA342986968640B CRC64;

Query Match 95.0%; Score 3836; DB 16; Length 758;
Best Local Similarity 95.7%; Pred. No. 3.1e-268;
Matches 728; Conservative 9; Mismatches 18; Indels 6; Gaps 3;

QY 1 MAQTTLKPIVLSILLINTPLLSQAHTQEQSGLETVSVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MAQTTLKPIVLSILLINTPLLSQAHTQEQSGLETVSVVGKSRPRATSGLLHTSTASDKI 60
QY 61 ISGDTLRKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADF 120
Db 61 ISGDTLRKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADF 120
QY 121 SPDHAIMVDTALSQQVEILRGVPTLLYSYSGNVAGLVDVADGKIPEKMPENGVSGLGLRL 180
Db 121 SPDHAIMVDTALSQQVEILRGVPTLLYSYSGNVAGLVDVADGKIPEKMPENGVSGLGLRL 180
QY 181 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVPRYRNKRLPSPRFANGQHRA 240
Db 181 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVPRYRNKRLPSPRFANGQHRA 240
QY 241 V-LGW--RKRFRYRYSRRDOYGLPAHSHVEDDCHADIWOKSLINKYLOLYPHLLTE 297
Db 238 IGLSWVGKGFAGYSDRRDQYGLPAHSHVEDDCHADIWOKSLINKYLOLYPHLLTE 297
QY 298 EDVDYDNPGLSCGFFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 357
Db 298 EDIDYDNPGLSCGFFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 357
QY 358 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRKSGWGQYLGQKSSALSATSEAVK 417
Db 358 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRKSGWGQYLGQKSSALSATSEAVK 417
QY 418 QPMLLDNKVQHSYFFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYKOPLPDL 477
Db 418 QPMLLDNKVQHSYFFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYKOPLPDL 477
QY 478 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
Db 478 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
QY 538 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQ 597
Db 538 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVRYNQ 597
QY 598 SGADFYGAEGEYIFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFFAODDQAP 657
Db 598 SGADFYGAEGEYIFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFFAODDQAP 657
QY 658 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANYRNTRYG 717
Db 658 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANYRNTRYG 717
QY 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758

RESULT 3
Q9AER9

Q9AER9	PRELIMINARY;	PRT;	766 AA.	
Q9AER9	PRELIMINARY;	PRT;	809 AA.	
01-JUN-2001	(TREMELrel. 17, Created)			
01-JUN-2001	(TREMELrel. 17, Last sequence update)			
01-MAR-2003	(TREMELrel. 23, Last annotation update)			
DE	Iron-regulated outer membrane protein.			
IRP.				
OS	Pasteurella haemolytica.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Mannheimia.			
OX	NCBI_TaxID=75985;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Lo R.Y., Graham M.R.;			
RT	"Putative TonB dependent receptor of Mannheimia haemolytica.";			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).			
DR	EMBL; AY028475; AAK29743.1; -			
DR	InterPro; IPR000531; TonB_boxC.			
DR	Pfam; PF00593; TonB_dep Rec; 1.			
DR	PROSITE; PS00430; TONB DEPENDENT REC 1; 1.			
KW	Membrane; Outer membrane; Receptor; TonB box.			
SQ	SEQUENCE 766 AA; 86440 MW; 7D76FAE05D7C5619 CRC64;			
Query Match	42.7%; Score 1724; DB 2; Length 766;			
Best Local Similarity	44.6%; Pred. No. 1.le-115;			
Matches	346; Conservative 128; Mismatches 260; Indels 42; Gaps 12;			
QY	9 IVLSILLINTPLAQAHETEQQSGVLETVVVGKSPRATSGLLHTS-TASOKIISGDTLR 67			
Db	7 LTVSILLALISG-VGYANE-----ISLETIIVDGNTPSYKGLLGGELNSVVDKLNK 60			
QY	68 QKAVNLGDALDGVPGHAGSYGGASAPVIRGQGRIRKVLNHHGETGMDAFSPDHIM 127			
Db	61 QGSITLGNALSGBLGHSHSQFGGASTPIIRGQESKRAKILQNGENLDMGMSPDHATV 120			
QY	128 VTALSOQVEILLRGPVTLVSSNGVAGLVADGKIPEKMPENGVSGLGLRLSSGNLEK 187			
Db	121 VVALAKLEILLRGPVTLVSSNGVAGLVADGKIPEKMPENGVSGLGLRLSSGNLEK 180			
QY	188 LTSGGINIGLKNFVLHTLEGLYKSGDYAVPRY----RNKRLPDSPPRFANGQHRVILG 243			
Db	181 LTVAGSTFALGNHALRVQGMYNKASEYVAPFTIEGKPYHRVPSDVQSQTG--TVSL 238			
QY	244 W--RKFRYRTYSDRRDQGLPAHSHYDDCHADIIWQSLINKRYLQLYPHLLTEEDVD 301			
Db	239 WIGERHGLIATYDRDRDKYGLTHYKDYHTISIRQAVMPAKGYLRFPYPLABEGDID 298			
QY	302 YONPGLSC-----GFHDDDAHAHNGKQWIDLRNRYELRAEWKQPPFGFEALRVH 354			
Db	299 YANPGLRLHATHIPGSHYQDTHET--GKPIDMHSKRYDIDGSLQNLPLFGFEAKIS 355			
QY	355 LNENDYHDEKAGDAVENFNQTONARIELRHQPIGRILKSGWGVYLGQKSSALSATSE 414			
Db	356 ANYVDYHDEKQGRYENFKNGKRLRFLVHKWKGLKGAIGVQYTNQSTALALEAS 415			
QY	415 AV-----KQPLNDNVQVHSYFPGVQANWNTFTLEGGVRVEKQKASIRYDKALIDRENY 469			
Db	416 RAAKVNKQPLNNPKTKLWSLPAIERNLNGDFTFELSGAERQKIAMDYDKLIDRLIG 475			
QY	470 YKQPLDLGAHROQTARSPALSGNWYTPPHQKSLTASHQERLPSTQELVYAHGKHVATNIF 529			
Db	476 FNTMPENLPHDKGYSYFATHWYFAPNPKLTNAHQERLPNAQELVYAHGKHIALNAF 535			
QY	530 EVGNKHLNERSNIIELALGYEDRWQYNLALVRFNGVYIAQTLNDGRGPKSIEDDSE 589			
Db	536 EAGNKLKERSNQIELSLIAYGDKWDYKLNLYHTHYGNYIYPLTNDNRGPKSFTDEYN 595			
QY	590 MKLVRYNQSGADFYAGBEIYFXTPTRYIGVSGDYVRGKLNKLPSP-----PQREDA 642			
Db	596 LKNRYIQGEARPSGAEGEIGLYFTPNYRLAVGQDVVRGKLNLPNIAMSYNIWTGEVDX 655			
QY	643 YGNRPFLAQDDQNAVRPAARLGFHLKASLTDRIDANLIDYRVRFAQNKLYARTETTPGH 702			
Db	656 WASQP-----DISAPRIPPLRLGARNADFNLMWSGMLRYRFAQKQVSKYEQVTPGH 710			
QY	703 MLNLGANYRNTYRGWNVVYKADNLNLSVYAHSSFLSDTPQMGSRFTGGVNVKF 758			
Db	711 QVNLGVTSNHFNQTEYQVFLKVDNLNKNQYHASYLPHIPQMGGRNAMLGNWISF 766			
RESULT 4				
Q9CLW8	PRELIMINARY;	PRT;	809 AA.	
AC	Q9CLW8, 2001 (TREMELrel. 17, Created)			
DT	01-JUN-2001 (TREMELrel. 17, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	Hypothetical protein PM1081.			
GN	PM1081.			
OS	Pasteurella multocida.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Pasteurella.			
OX	NCBI_TaxID=747;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	STRAIN=PM70.			
RA	MEDLINE=21145866; PubMed=11248100;			
RA	May B.J., Zhang Q., Li L.B., Paustian M.L., Whittam T.S., Kapur V.;			
RL	"Complete genomic sequence of Pasteurella multocida PM70.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).			
DR	EMBL; AE006149; AAK03165.1; -			
DR	InterPro; IPR000531; TonB_boxC.			
DR	Pfam; PF00593; TonB_dep Rec; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 809 AA; 90903 MW; 23026B7DFCBF14DB CRC64;			
Query Match	37.0%; Score 1492.5; DB 16; Length 809;			
Best Local Similarity	42.9%; Pred. No. 6.6e-99;			
Matches	333; Conservative 120; Mismatches 271; Indels 53; Gaps 13;			
QY	27 TEQSVGLETVTVGKSPRATSGLLHTSTASDKIISGDTLRKAVNLGDALDGVPGH 86			
Db	41 TSEHIEIFVNTLIBESREGAPLGGRLMASEKIIIPAYSKQSGNLGDSLSSELGHAS 100			
QY	87 QYGGASAPVIRGOTGRIRKVLNHHGETGMDAFSPDHAIMVDTALSOQVEILLRGPVTL 146			
Db	101 QFGGASAPVIRGEGAKRIKVLSCNETLDMASPDHAVAUSGLLAKKVEILLAGANTLL 160			
QY	147 YSSGNVAGLVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGINIGLKNFVLHTE 206			
Db	161 YSSGNAAGVNVVNDKIP-TAEVVGVEGVLRTGSADNERLVNVALDVGSLKFHALHE 219			
QY	207 GLYKSGDYAVPRY----NLKRLPDSPPRFANGQHRV-LGM--RKFRYRTYSDRRDQ 259			
Db	220 GLHKAGDYRTPSYQSGSTHKLANS---FVDRSGSVGLSWVGDKYLGVAYSQRDK 276			
QY	260 YGLPAHSHYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDVNDPGLSC-----GFHDD 314			
Db	277 YGLPAHSHLYDEYVMVLLSDAHRKPYLKHYPFMEETDIDYNNPIDCICKKSHSHGH 336			
QY	315 DDAHAHNGK-----PWIDLRNRYELRAEWKQPPFGFEALRVHLNNDYHH 362			
Db	337 LCNIGHAHNGQSHDHHADPHIALNTQRMDLRGWKNPKVGLDKVRFSTAKVGSRH 396			
QY	363 DEKAGDAVENFNQTONARIELRHQPIGRILKSGWGVYLGQKSSAL-SATSEAVKQPM 421			
Db	397 DEKSGALSNSFKNGYSARVEFLHQPIAGVSLGLSHVYQDSYALDMHTLEYKQNL 456			
QY	422 LDNVQVHSYFPGVQANWNTFTLEGGVRVEKQKASIRYDKALIDRENYKQPLDL-GAH 480			
Db	457 SDHTTAQOQLFMEHVELGKQWFDIGRVEKQRIAMKYHFNVPKDE---QPPELTRPH 512			
QY	481 RQTARSFALSGNWYFTPHQKLSLTASHOERLPSTQELVYAHGKHVATNTEFGNKHNER 540			
Db	513 KSKAYSALSNYQLNEHQHFNMLVSHQERLPNAQELVYAHGKHVATNTEFGNKNUTYER 572			

```

QY 541 SNNIELALGVEGRWYNNIALYRNFRGNYIAOTLN-----DGRGPKSIEDSEMK 591
Db 573 SNNVELGWGTGKGLGKLSGGYQQFSNYIYAAILNNKTSCEPWRPNRSLRSLSDYPLR 632
QY 592 LVRYNOSGADFYGAEGEIVFKPFRVRIYVSGDYVRGLKNPLPSLP-----GREDAY 643
Db 633 LYRYNQAKIYGLAEVSVQISSTHSVIFG DYVRGKLKOLPSLPFGYKYIYNNYDMV 692
QY 644 GNRP--FIAODQNAPRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETTPGH 701
Db 693 GVQPTGWEKQPDGNAPRMSPMRLGIXKNWAFDNGISFNTQLXRVFAQNKVARLETPTKG 752
QY 702 HMLNGLNRYNRTRYGEMWVYKADNLLNQSVVAHSSFLSDTPQMGRSFTGGVNVKPF 758
Db 753 TMLNLGMSYDGKMGNNETTLFANVNVNLSRVNHTSFLSYIFQSGLGLNVMGNVFXF 809

RESULT 5
Q9CL20 PRELIMINARY; PRT; 805 AA.
ID Q9CL20 AC Q9CL20
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein PM1428.
GN PM1428.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006179; AAK03512.1; -
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 805 AA; 90850 MW; 799A05B9E4D10509 CRC64;

Query Match 33.8%; Score 1364; DB 16; Length 805;
Best Local Similarity 37.3%; Pred. No. 1.3e-89;
Matches 294; Conservative 136; Mismatches 291; Indels 68; Gaps 10;

QY 26 ETEQSVGLTETVTVGKSRPRATSGLLHTSTASDKITSGDTRQKAVNLGDALDGVPGIHA 85
Db 29 EKKQIVFDEISVESKGAFRSDPLSGLPKQNDILVSKQKLTGSGSTLGNALAGELSVHS 88
QY 86 SQYGGASAPVIRGTGRIKVLNHHGTGDMADFPDHAIMVDTALSOQVEILRGVPTL 145
Db 89 NQFGGSSAPVVRGQGVRLKTLQNGSDVIDMSQLSPDHAIGVDTLLAEQVEIVRGASTL 148
QY 146 LYSNGNVLVDVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGGINIGLGNFVLHT 205
Db 149 LYANASAPGVINVDKRIPTQPKGYEVDFNTRYNTNSHEKLVTAALFTGLGHIALRV 208
QY 206 EGLYRKSGDYAVPRY---RNLRKLPDPSRPFANGQHRVAVLGRKRFYRTYSDRRDQVGL 262
Db 209 EELLRGNNHYVPAFLDKTLNVPDTQNKTKSGNYGVAFTGERGYGVFAYNLRREKYGL 268
QY 263 PAHSHEYDDCHADIIWQSLINKRLVQLYPLHLLTTEEDVDYDNPGLSCGF-HDDDDAHAHA 321
Db 269 PGNHKLDSAAH-IWGNVNRNDYLYGLYPLHLM--HDTDLVNTPHCGSNHMDGKSHN 325
QY 322 HN-----GKPMIDLNKREYLRABWKQFPFGFEALRVHLNENDYHHDKAGDAVEN 372
Db 326 HPYGHDDHDSIAGPLIDSYAKRYDIFRAEVKQPMKAIEKILSYSETRYKHKDEKGNIAVN 385
QY 373 PFNQTONARIELRHQPIGRLGSGWGVYLGQKSSA----LSATSEAVKQP----- 419

```

```

Db 386 LFKNGYNLRVIEIHTPIAGLSGVIGAYQYQOTQSSANIPRIAPCSNNASDPCHKKKORDP 445
QY 420 -----MLLDNKVQHYSPFGVQAWNDFTLEGGVYRVEKQKASIRYDKALI----- 464
Db 446 SKITGDKRKSALLIENTOSQMSFAIEQLRWQDFLFEIGVTRTEKQRIIDIEVRAWLEFKVK 505
QY 465 -----DRENYI-----KQPLPDLAGHARQOTARSFALSNGWYFTPOHKLSTASHOE 509
Db 506 KLEGCDFNSFFYSPSGCRQGSYPAPDFASVYDRATSYSGAISNNMTDPYTLSTLYSHNE 565
QY 510 RLPSTQELIYAGKGVAVNTTFVGNKHLNKNERSNNMIELALGHEGDRWQYNLALYRNGNY 569
Db 566 RHPTMELYHGHKHLATVSEFHGNRLKKEVSDNWVEGLAYLGLDKLSYKNVYVYNDFKNR 625
QY 570 IYAOTLNDGRGPKSIEDSEMKLVYRQSGADFYGAEGEIVFKPFRVRIYVSGDYVRGR 629
Db 626 IFNQTLN-----KSGNLRLRYNQSKAKYIYGVGGRIDYALTPELHMGIFGDYVRGK 676
QY 630 LKNLPSLPGREDAYGNRPFIAQDDQNAPRVPAARLGFHLKASLTDRIDANLDYRVFAQN 689
Db 677 LYDLPTVYRVHDVANSLEPVTQPDQADAPRVPEMLRIGFRVNMMENTESTLSLEYTVYVQK 736
QY 690 KLARYETTPGHMLNGLNRYNRTRYGEMWVYKADNLLNQSVVAHSSFLSDTPQMGRS 749
Db 737 KVAPLENQTAAYSLNLIQVDSYRQIAGVNYQLFVQANNVLRKRYSHTSFLPFVPMGRN 796
QY 750 FTGGVNVKPF 758
Db 797 VTGLGNIHF 805

RESULT 6
Q9CMR7 PRELIMINARY; PRT; 925 AA.
ID Q9CMR7 AC Q9CMR7
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein PM0745.
GN PM0745.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006112; AAK02829.1; -
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 925 AA; 105081 MW; F46A9F52B59145E3 CRC64;

Query Match 32.4%; Score 1307; DB 16; Length 925;
Best Local Similarity 33.7%; Pred. No. 2.1e-85;
Matches 292; Conservative 149; Mismatches 278; Indels 148; Gaps 14;

QY 29 QSVGLETVTVGKSRPRATSGLLHTSTASDKITSGDTRQKAVNLGDALDGVPGIHA 88
Db 70 QTEMLDEIVVSGAQPHLAGSAIEHYQAIISNOVKKERLQKQATLGNALAGELGVHSNFF 129
QY 89 GGASAPVIRGTGRIKVLNHHGTGDMADFPDHAIMVDTALSOQVEILRGVPTL 148
Db 130 GGASAPIIRGQGVRIKILQNGLDYVDMASAIKSPDHAIVADSLAEQVEILRGASTL 189
QY 149 SGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGGINIGLGNFVLHTGL 208
Db 190 SASAAGVNVVDKRIETAVFEKGYEGEITFRFDTSQBSTGTAGITFRHLPHLALLEGSL 249

```

```

QY 209 YKSGDYAVPRYN---LKLRLPDSPPRPFANGHRAVLGWRK--RFRRTYSDRDYGL 262
Db 250 KRYSTHYRVPKSGETIRYLPDSHNRSQVG--TIGVSWIKDQSYGLVGSYSERRDRYGL 307
QY 263 FAHSHEYDDCHADIIWQ--KSLINKRYLOLYPHLLTEEDVDYDNFGLSCGFHDDDDAHAAH 320
Db 308 PGNHKKYDRCKSHVVDDEARPELGGKYLTPYPLHADDTDIVAHLDGCGTGGIDNDFSHSH 367
QY 321 AH-----NGKPWIDLNRKRYELRAEWKOPFGFEALRVHLNRNDY-HHDEKAGDA- 369
Db 368 DHFPGEHDSHGGPWLRSRFDLRGQWDSPTAWLDKVGKFSFYADYIHYEYHSGQAG 427
QY 370 -----VENFFNNTQONARIELRHOPIGLKGSGWGYLQGS 406
Db 428 TDKPRDSFIERERKAEKVRKAAGIYKNSGVNGKLEFYHTPIVGLSGVFGVQVSEYKT 487
QY 407 SALSATSEAVK-OPMLLDNKVQHYGFFGVEQANWONFTLEGVVRVEKQKASIRYDKALID 465
Db 488 AILAPLGSIGIKHQHLLVPNTQOAGFFAVENVVDDDFIEVGARVDKQRIPIKYDQHVLN 547
QY 466 RENYYKQPLPDLGAHQRTARSFALSGNWFYTPQHKLSLTASHOERLIPSTQELYAHGKHVA 525
Db 548 AHKKGSDHPDPLSTHKEKAVSLGSDVWDLFHNRYRIGLTLRNERLPTPMELYHYHQHLLA 607
QY 526 TMTFVGNKHLNKRNSNIELALGYEGDRQWQYNLALYRNRFGNYIAOYTLNDGRGPKSTE 585
Db 608 TNSFHEGNKLRKESNNIELGFAHTDKWDYKLSLYQNKFRNYLYNEDL--ARYGNAP- 664
QY 586 DSEMKLVRYNSGADFYGAGEIYFKPTPRYRIGVSGDYVRGRLKNLPSLFGREDAYG- 644
Db 665 -----LRRYTQARAKFHGIEAELNRPPTPDYQVTLFGDYVRGRLFDLPEQYQRFYQY 718
QY 645 -----NRPF-----IAQDDQNAVRPAPARLGFHLKASITDRIDANLDYRVF 686
Db 719 IAYDEGLQANWEKQPIYRIGIKRNERDAPRVPARLGMRLSGNVTEHLSFFADYTYVF 778
QY 687 AQNKLA-----RY----- 694
Db 779 SQOKTASSLSIKPPRALEASDFIDDTGENILLKGIDRKYNRTQOPTADLSASDKADPE 838
QY 695 -----FRTPGHMLNIGANYRNRTRYGEMWNYVKADNLLNQ 731
Db 839 SALEAIRLEENKELPAKTEKIQEDPSKGHLLNIGVNYQRIQIHLDSYSGVSVANNVQ 898
QY 732 SYVHSSFLSDPTQMGSRSTGCVNWK 758
Db 899 RVYVHTSYLPVQPMGRNVLNFGVKF 925

RESULT 7
ID Q8P7F3
AC Q8P7F3 PRELIMINARY; PRT; 688 AA.
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Outer membrane hemin receptor.
GN PHUR OR XCC2658.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannavaro F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

```

```

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT *Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.*;
RL Nature 417:459-463(2002).
DR EMBL: AB012378; AM41930.1; --
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 688 AA; 74262 MW; 2E47F6BB47F6429B CRC64;

Query Match 18.5%; Score 745; DB 16; Length 688;
Best Local Similarity 27.5%; Pred. No. 5e-45;
Matches 219; Conservative 123; Mismatches 301; Indels 152; Gaps 21;

QY 1 MAQTTLKPIVLISILLINTPLLAQA-----HETQSVG-----LETVT 37
Db 9 LRERSTLTALASVLI---PALAMADTVPADPQTPPSSDTRHDATPPSGRIHKDVKV- 64
QY 38 VVGKSRPRATSGLLHTSTASDKIISGDTLRQ-KAVNLGDALDGVPGIHASQYGGGASAPV 96
Db 65 VVTASPLRDAAGELSRPV---EVLAGERLDEVRASSLGETVASLPGVQSSNFGPGVGRPI 121
QY 97 IRGQTRRIKVLNHHGEIGTDMADFDPDHAIMVDYDTALSSQVEILRGVPTVLLYSNGNVAGLV 156
Db 122 IRLGLDPRVAVLRDLGLSTQDVSTVSDHSPAIEFLANQIEVLKGPSTLLYSGAIGGVV 181
QY 157 DVADGKIPEKMPENGVSGLRLSSGNLEKLTSGGINIGLGNKFNVLHTEGLYRKSQY 215
Db 182 NVVDGRIAE-TPVDGFSGEAEVRFQGGKDGNTDMFRVDAGNGSGLSIHADGVYRNQDY 240
QY 216 AVPRYRNLRKLPDPSRPFANGQHRVAVLGMWRKFPYRTYSRRDQYGLPAHSHEYDDCHAD 275
Db 241 DTPQGRQANSFLDS---KVGSGISAGLADMGVGLSASRFRDNYGNPGE----- 286
QY 276 IIWQSLINKRYLOLYPHLLTEEDVDYDNFGLSCGFHDDDDAHAAHNGKPMIDLNKRY 335
Db 287 -----PGDLSIGERGS-----LKLQDDRY 306
QY 336 ELRAEWKOPFGFEALRVHLNRNDYHDEKAGDAVENFFNNTQONARIELRHQPIGRKLG 395
Db 307 DLKGGLTDPWGDGSAIRYSFGHTDYAHTFEAGAEVGTVTKRANEGRVEASTFGGQWQT 366
QY 396 SWGVQYLQKKSALSATSEAV-KQPMILDNKKVQHYSGFFGVEQANWONFTLEGVVRVEKOK 454
Db 367 AFGLO-----GSDTTFQAVGEESFVPTKDTRSLSGLFGVARNSDRVTAEVGARVDVKV 419
QY 455 ASIRYDKALIDRENYKQPLDGLAHRQ-TARSAFALSGNWFYTPQHKLSLTASHOERLPS 513
Db 420 YS-----TDIGVDRDFTPTFSASAGFRNEQWRLTANLIDHARAPA 461
QY 514 TQELYAHGKHVATNTFEVGNKHLNKRNSNIELALGYEGDRQWQYNLALYRNRFGNYV- 571
Db 462 EEELFANGPHIATLAFEGIDANILKTEKANQALGNLFQNDWDVDAKIAAYNRINDFVYIV 521
QY 572 ---AQTLDNDRGPKSIEDDSEMKLVRYNSQADFYGAGEIYF-----KPTPRYRIGVSGD 624
Db 522 DTGGQWFNE-----EDNDFLPIRQWTAQDAIFHGPEGEATFHLANNDTGAWDLRVFGD 574
QY 625 YVRGLKNLPSLPGREDAYGNRPFIADQDQNAVRPAPARLGFHLKASITDRIDANLDYR 684
Db 575 TVQARLK-----DGGNLPRIAPGRVGAEMRW-N-ADAWRASLGAI 613
QY 685 VFAQNKLARVETRTPGHMLNIGANYRNRTRYGEMWNYVKADNLLNQSIVYHSSFLSDTP 744
Db 614 TMKQKAVVNETATDGYTWDVAHLAYHMDRGANAVEVFLDGNLNTNQDARVHTSFLKDDV 673

```

```

QY 745 QM-GSFTGCVNKF 758
DB 674 MLPGRNASFGVRLFF 688

RESULT 8
OSPIR6
ID Q8PIR6 PRELIMINARY; PRT; 687 AA.
AC Q8PIR6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane hemin receptor.
GN PHUR OR XAC2829.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Perro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavar F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Perro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AB011924; AAC37674.1; -.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep.Rec; 1.
KW Receptor; Complete Proteome.
SQ SEQUENCE 687 AA; 74280 MW; A0B719C4ED4F6D1F CRC64;

Query Match 17.8%; Score 718.5; DB 16; Length 687;
Best Local Similarity 27.1%; Pred. No. 4.1e-43;
Matches 213; Conservative 122; Mismatches 301; Indels 149; Gaps 21;

QY 11 LGLINTPLLAQA-----HETEQSVG-----LETVTVGKSRPRAT 47
DB 15 LALASLLTPALMAEDAPVAADQGPSPSDTHDASPSSGRHIXDLDKV-VTASPLDA 73
QY 48 SGLLHTSTASDKIISGDTLQ-KAVNLGDALDVPQIHASQYGGASAPVIRGTGRK 106
DB 74 AGELSRPV---EVLAGERLDEVRSSLSGETVASLPGVQSSNPGVGRPIIRGLDGPVA 130
QY 107 VLNHGEGTDMADFSPDHAIMVDVTALSQVETILRGVTLVLYSSGNVAGLVADVADKIPK 166
DB 131 VLRDGLSTQDVTSTVQSDHSPAFEPFLAQIIEVLRKGFSTLLYSGSAGVGVNVVGRIAE- 189
QY 167 MPENGVSGBLGRLLSSGNLEKLTSGGINIGLKKFVLHTEGLYKRSQDGYAVPRYRNLKR 225
DB 190 TPVDGFSGRAEVRFDGDKDGTDMFRVDAGNSGSLSHADGVYRNQNDYTPKGRQLNS 249
QY 226 LPDSPRRFANGQHRVILGWKRFYRTYSDRRDQGLPAHSHEYDDCHADIIWQKSLINK 285
DB 250 WYDS-----KVGSGTASLSDGWDGVLGSASRFRDNYGNPGE----- 285
QY 286 RYLQLYPHLLTEEDVDYDNPGLSCGPHDDDDAHAAHAKPWIDLNRKRYELRAEWKOPF 345

```


RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira M.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
RL Nature 406:151-159(2000).
DR EMBL: AE003890; AAF83194.1; --
DR InterPro: IPR000566; Lipocln_cytFABP.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Reg; 1.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Complete proteome.
SQ SEQUENCE 681 AA; 74084 MW; E7FFDF36F1CEE5ED CRC64;

Query Match 17.28; Score 694.5; DB 16; Length 681;
Best Local Similarity 27.4%; Pred. No. 2.2e-41;
Matches 220; Conservative 115; Mismatches 294; Indels 173; Gaps 22;

QY 1 MAQTTLKPIVLSILLINPTLLAQA-----HETQSVGLETVTVVG----- 40
DB 9 LPRTVFLAVSSLL--TPALAMADVGPPTDSRHKFTTEVTRHLKDLDAVVVTAIPLRD 66
QY 41 ----KSPRATSGLLHTSTASDKLIISGTLRO-KAVNLGALDGVGIIHASQVGGASAP 95
DB 67 SYDLSRPVA-----LLAGERLDEVRASSIGETAVLPVQVSNFPGVGRP 113
QY 96 VIRQGTGRRIKVNLHGHETGMDAFSPDHAIMVDITALSQOVELRGPTVTLTSSGNVAGL 155
DB 114 IIRGLDGPRIAVLNGLSSQDVSTVSDHSPAVEPFLANQIEVLKGPSTLLYSGAIGGV 173
QY 156 VDVAADGIPEKMPENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLTTEGLYKSGD 214
DB 174 VNIVDGRIAE-APYGVGFNGRAEMRLDGGDKHGNTNMFRI DAGNGSALSVAHDGYRNEKD 232
QY 215 YAVPRYRNLKLPDSRRFANGQRAVLGWRKRYRRTYSDRDQYGLPAHSHEYDDCHA 274
DB 233 YDTPKGQVNSFIDT----KSGSWGASFSGDGFGFGLSVARFHDYSY----- 275
QY 275 DIIWQKSLINKRYLYPHLLITEEDVDYDNPGLSCGPHDDDDAHAAHNGKPIDLNRK 334
DB 276 ----NFG-----EPGDVPAGDRGSLRLRHQDR 298
QY 335 YELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNOFNQONARIELRHQPTGRLK 394
DB 299 YDLKAGLTDPMGEASALRPSLCHTQYDIEPEGNEVGTTFGRASEGRVSAFAFGGWR 358
QY 395 GSWGV-----QYLGQKSSALSATSEAVKQPMLLDNKVOHYSFPQVEQANWNTFLEGG 447
DB 359 TAFGVQGGDSTFOALGEESEFVPEKTNKSI-----GVFGLAHNTFGLFQAEFG 405
QY 448 VRVEKQKASIRYDKALIDRENVYKQPLDPLGAHQRTARSFALSGNWWYFTPHKLSLTASH 507
DB 406 ARGDQ-----VKYDITDNGVTRNYH-----PGSLAFSGDLALSQKWLRLTNVDH 448
QY 508 QERLPSTQELYAHGKHVATNTFEVGNKHLNKRNNIELALGYEGDRQWYNLALYRNRF 567
DB 449 AERAPVEEELFAKGPHIATLAYEVRGADLKEKANAQELGLVFRDWSDAKVSTYSRYG 508
QY 568 NYIYAQTILNDGRGPKSIDDESEMILV--RYNSGADFYGAGEIYF-----KPIPRYRIGV 621
DB 509 NFYI---LVD-TGTSFWDDQORDLFRVQWSQANAI FHGIEGEATFHLAKNTSGSDLRV 564
QY 622 SGDYVRGLKXNLPSPGREDAYGNRPFFIAODCONAPRVAPARLGFHLKASITDRIDA--- 678
DB 565 FGDTVSGRLKN-----GGNLPRIVPARYGAELRWE-----DAGWR 599
QY 679 -NLDYVRPAQNKLYARTPTFGHMLMLGANYRNTRYGWNVYKADNLLNOSVVAHS 737
DB 600 TSLSAKRYEKQNKVAVNETPTAGYTWDAHLAYIDVDSIAWEVFFDGNLNTNRDARVHT 659
QY 738 SELSDTPQM-GRSFTGGVNVKF 758

DB 660 SFLKDDVMLAGRNYTAGLEMPFF 681

RESULT 10

Q915F7 PRELIMINARY; PRT; 687 AA.

ID AC Q915F7; 687 AA.

AC Q915F7; 687 AA.

DT 01-MAR-2001 (TRENBLrel. 16, Created)

DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 23, Last annotation update)

DE Hypothetical protein PA0781.

GN PA0781.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Brody R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saiter M.H., Hancock R.B.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004513; AAG04170.1; --
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Reg; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 687 AA; 74205 MW; FBEC6ACD94B8FDD CRC64;

Query Match 16.8%; Score 679; DB 16; Length 687;
Best Local Similarity 27.4%; Pred. No. 2.9e-40;
Matches 212; Conservative 121; Mismatches 319; Indels 122; Gaps 25;

QY 6 LKPIVLSILLINPTLL-AAAAHETQSVGLETVTVVVGKSRPRATSGLLHTSTASDKLIISGD 64
DB 15 LTPALACILVSGETIGADGRPSLPQSVITANPLGNESPATPS-----SVLEGD 64
QY 65 --TLQKAVNLGDALDGVGIIHASQVGGASAPVIRGOTGRRIKVNLHGHETGMDAFSP 122
DB 65 EHTLQKQ-SLGETNLGLPGVSTTFPGASRPVIRGMDGDIRLLRNGVGLDASSLSY 123
QY 123 DHAIMVDITALSQOVELRGPTVTLTSSGNVAGLVADGKIPEKMPENGVSGLGLRLSS 182
DB 124 DHAVPEDPNSVERLEVRGPAALYGGNAIGGVNSFDNRIPSE-PVDGIHSGSLRYGG 182
QY 183 GNLEKLTSGGINIGLGNFVLTTEGLYKSGDYAVPRYRNLKLPDSRRP--ANGQHRA 240
DB 183 ADTTRSRSGALEAGDG-NFALHVDAAAREFDNRIPGYAHSR----QRQIDGTGRHV 237
QY 241 VLGWKRKRYRRTYSDRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLYPHLLITEEDV 300
DB 238 -----QNSDGRQDGGAVGSGYHWEHGYAGLSY-----SGYDS 269
QY 301 DYDNPGLSCGPHDDDDAHAAHNGKPIDLNRKRYELRAEWKQPPGFEALRVHLNRNDY 360
DB 270 NYGSPA-----EDVNR-----LKMQODRYAFASEINDLEGPTSLKDAAYTKY 313
QY 361 HHDEKAGDAVENFNNOFNQONARIELRHQPIGRKLSGWVQYLGQKSSALSATSEAVKQPM 420
DB 314 EHKELTGTGTTFKNEGYEGRIEAEHRPLGVLNGVGAQFANSRFSALG-----EAF 367
QY 421 LLDNKVQHYSPFGVQANW---DNFTLEGVREYKQKASIRYDKALIDRENVYKQPLD 477
DB 368 VPHETDTSAALEAE--WKLSDRLDLSFGARLE----HTRVDPDAKGNRFAEND---- 417
QY 478 GAHQRTARSFALSGNWWYFTPHKLSLTASHOSELSTQELYAHGKHVATNTFEVGNKHLN 537


```

Db 418 GSQSTTGSSTGAVYKLTPIWSLAATLSYTERAPTFYELANGPHAATGYEVDADAD 477
Qy 538 KERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLDNGRGPK-----SIBDDSEMK 591
Db 478 KEKAVSTDLALRFDFGVHKGSGVGFYSRFSNYI--GLLASGRHNEEGVVAAGDDEALP 535
Qy 592 LVRYNQSGADYFGAEGE-----LYPKPTPRYRIGYSGDYVGRGLKNLPSLPREDAYGNRP 647
Db 536 EYLYSGVWRADYFGEAQDRHLLHSPYCNFDELSGDYTRAKNKD-----580
Qy 648 FIAQDDQNAQVPAARLGFHLKASITDRIDANLDYRVFAQNKLARVETRTPGHMLNLG 707
Db 581 ----TGEPLPRLPLRLNTALIWEL--QQQARVDVEHAASQHRVPEELSDGY--TTLG 633
Qy 708 ANYRNTRYGWNW--YKADNLLNQSVYAHSSFLSD--TPQGRSFTGGVNVKF 758
Db 634 ASLGYNFDLGSERWLAFLVKGNTLNTQTVRYASSILDRVPAAGRGIEAGVKVAF 687

RESULT 11
Q9ABL3
ID Q9ABL3 PRELIMINARY; PRT; 687 AA.
AC Q9ABL3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE TonB-dependent receptor.
GN CC0214.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Winn M.B., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RC Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005695; AA22201.1; -.
DR TIGR; CC0214; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 687 AA; 72786 MW; 29415A726A3BF753 CRC64;

Query Match 15.8%; Score 636; DB 16; Length 687;
Best Local Similarity 26.5%; Pred. No. 3.7e-37;
Matches 206; Conservative 107; Mismatches 237; Indels 168; Gaps 23;

Qy 22 AQAHTEQSGVLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLRQKA-VNIGDALDGV 80
Db 37 ASAQDAEVSKVIITAPYGVY-----ADALTASVAVLDRTDLDAAPKGLGDLALG 88
Qy 81 PGHIAQYGGGASAPVIRGTGTRIKVNLNHGFTGDMADFPDHAIMVDTALTSQQVEILR 140
Db 89 PGVRSSTTFGAGSRPVVRGLAGPRVQVLTNGVGQIDASALSPDHQVATDPGEAERIEVL 148
Qy 141 GPVTLLYSSGNVAGLVADVAGKTPKPNPENGVSGLRLSSGNLEKLTSGGINIGLGN 200
Db 149 GPAAIAYGSAIGVGNVNIIDRISTQQPIDGMSGRLLASRGTDGSDSALSAGVHATVGP- 207
Qy 201 FVLHTEGLYKSGDYAVPRYRNLRKLPDSPRFA-----NGQHR-----AV 241
Db 208 MVLTLDAKLESKDYKIPVY-----PESARQALAEGETAEGAGRLSENSAVDLETFGAG 261
Qy 242 LGW--RKRFRYRTYSRRDQYGLPAHSHVEDDCHADIWQKSLINKRYLQYPLHLLTEED 299

```

```

Db 262 LSYVDKRGFGVMSIKRTDSTDYGVPCHAHE-----290
Qy 300 VDYDNPGLSCGFHDDDAHAHANGKPMWIDLRNKRXYELRAEMKQFPFGFEALRVHLNRND 359
Db 291 -----HGEABAGHEESAVTIGUKQTRIDLRGEBYDADLGGFAKVRFGSGHAD 338
Qy 360 YHDEKAGDAVENFNQNTQNAIRIELRHQPTIGRLKSGMWGVYLGQKXSALSATSSEAVKQP 419
Db 339 YTHTEFGDAGVTRFTSDGYEGRLVQTERGGWKAGVGVQALARNFPAIDGAEAVPK-- 396
Qy 420 MLLNKVQHYVFFGVQGANWDFLEGGVRVEKQKASIRYDKALIDRENYKQPLDGLA 479
Db 397 ----TKITEFGAFTQORLELDGYEGGLRITDRELDSLKGA--DFTN-----439
Qy 480 HRQTARGSFALSGN-----WYFTPOHKLSTASHQRLPSTQELYAHGKHVATNTFEVGNKH 535
Db 440 -----LGSVGAFWRPTTESFVGLALSRSRAPTESELFAGPHAATRGFEIGDAD 490
Qy 536 LNKERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLDNGRGPKSIEDDSEMKLVRY 595
Db 491 LKEETATSLAETHYGERVSGDLHLVIARYDGFIDLRPTGD-----EEDG-LAVRY 542
Qy 596 NQSGADYFGAEGEYIFKPTPRYRIGVSG-----DYVRGLKNLPSLPREDAYGN 645
Db 543 VQTDABEFGFEAEI-----AYRVWTDGQSRVNLHAGADFVRG-----579
Qy 646 RPFIAQDDQNAQ-RVPAARLGFHLKASITDRI--DANLDYRVFAQNKLARVETRTPGHHM 703
Db 580 -----SSDLGPPARIPYVS--LKASYQAPMWSGDVEVTRTGGQERVAQNELPTDGYTV 632
Qy 704 LNLGANYR--RNTRYGEMWYKADNLLNQSVYAHSSFLSD--TPQGRSFTGGVNVKF 758
Db 633 LNAATLAWKPAGNARV--RLFLDGRNLTNEVREHVSFLKDIAPSPGQVRAGATLRF 687

RESULT 12
Q8F1I6
ID Q8F1I6 PRELIMINARY; PRT; 777 AA.
AC Q8F1I6;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hemin receptor.
GN LA3149.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011477; AA050347.1; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 777 AA; 86615 MW; 03F985150B36589A CRC64;

Query Match 9.8%; Score 395.5; DB 16; Length 777;
Best Local Similarity 22.5%; Pred. No. 1.1e-19;
Matches 173; Conservative 121; Mismatches 305; Indels 169; Gaps 29;

Qy 28 EQSVGLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLR-QKAVNLGDALDGVPGIHAS 86
Db 105 ESSLDQSAINVTAKS---TNSDFLTAQPITVLSGRQLDRQGETAMSAINNTPGVSNL 160
Qy 87 QYGGGASAPVIRGTGTRIKVNLNHGFTGDMADFPDHAIMVDTALTSQQVEILRGPVTL 146
Db 161 TTSGSGTKPIIRLLTQGVRLVMT-DGIRQEBEQFGDDHTVELDSFNKIEIRPGSL 219
Qy 147 YSSGNVAGLVADVAGKIPKMPENGVSGELGLRLSSGNLEKLTSGGINIGLGNFVLHTE 206
Db 220 YGSDALGGVNVIRDKAP-----LSGE-----GIPKAGIFNS 252

```



```
QY 522 KHVATNTEVGNKHLNKNRSNNIELALGYEGDRWQYNLALYRNFPGNYYIAQTLDNRGP 581
D 502 SAGSVTY--GNPDLKAEHSNNFELGARYNGNWLIDSAVYYSKADYI--ASLICDGSIV 558
QY 582 KSIEDDESEMCLVRY--NQSADGFYCAEGEYFKPTPRYRIGVSGDYVRGRL----- 630
D 559 CNGNTNSRSRYYYDNDIDRAKTWGLE-----ISAENYNGWVFPYISGNLIRQYETST 612
QY 631 -----KNLPSLPGR-----EDAYGNRPPIAODDNA--PRVPA-ARLGFH 667
D 613 LKTTNTGEPAINGRIGLKHILVMQANILSDVFIARAASSAKDSDNGTETNPGWATLNFA 672
QY 668 LKASLTDRIDANLDYRV-PAQNKLA--RYET 696
D 673 VNTEF-----GNEDQYRINLALNLTDKRYRT 699

RESULT 16
Q8FGAO PRELIMINARY; PRT; 715 AA.
AC
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative outer membrane receptor for iron compound or colicin.
GN C2482
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016762; AAN80941.1; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 715 AA; 79177 MW; F4B7BE266775537B CRC64;
```

```
Query Match 6.6%; Score 264.5; DB 16; Length 715;
Best Local Similarity 22.5%; Pred. No. 2.7e-10;
Matches 169; Conservative 118; Mismatches 328; Indels 137; Gaps 38;

QY 11 LSTILLINTPLLAQHETEQQVGLTETVVGKSRPRATSGLLHTSTASDKIISGDTLRQKA 70
D 15 LSLCISIVTATSTSETKISNETLVTNTR---SASNLWSPATQIVDQOQLNST 71
QY 71 -VNLGDALDGVPGHIAHQY--GGASAFVIRGQTGRIKVLNHHGE--TCDMAFSPDHAIM 127
D 72 NASIADNLQDIPGVEITDNLISLAGFKQIRIRGEASSRVLILIDGQEVYQAGDNYGVCLL 131
QY 128 VDTALSOQVILRGVPVLLYSNGVAGLVDVADQKIPKMPKENGVSQELGLRLSSGNLEK 187
D 132 IDSALSERVVRVGPYSLVYSGAIGGINVFITKGGDKL-ASGVAVKAVNSATAGWES 190
QY 188 LTSGGINIGLKNFVLHTEGLYRKSQDYAVPRYENLKRLPDPSRPFANGQRAVLGHRKR 247
D 191 IAVQG-SIG--GFDYRINSYSDQNRDTPD-----GRLENT--NYRNNSGQVWLGVNSG 240
QY 248 FYRTVSDRDQYGLPAHSHEYDDCHADIWQKSLINKRYLLQLYPHLLTEE-----DVVD 302
D 241 NHR--FGLSLDRYLAQTYY-YEDPDGSY-----EAFSVKIPKLERKVGVFYDVTDV 289
QY 303 DNFGLSCGFHDDDAHAHANGKFWIDLNRKRYELRAEWKQPFPGFALRVHLNRNDYHH 362
D 290 DGYLK-----KIHFDYAEQTIQRFAN---EVKTTQVPVPMIQLATVH-NKTD-TH 337
```

```
QY 363 DEKAGDAV--ENFFNNOTQNIARIELRHQPIORLKGSGWGVQY-----LGQKSALSATSBAV 416
D 338 DKQYTOAVTLQSHFSLPANNELVT-----GAQYKQDRVSRSGMTSSKSLT 384
QY 417 KQPMLLDNKQVHYHFFGVEQA-----NW---DNFTLEGGRVVEKOKASIRYDKALID 465
D 385 G---FINKETRTSRYSSEQSTVSLFAQNDWRFAHDHWTWTWGV--QYWLSSKLTG--D 437
QY 466 RENTYKQPLDLAGHRTA--RSPALSNWYFTQHKLSLTASHQER--LPSTOLYAHG 521
D 438 GVSYTAGIISDTSLARESASHEMVTSTLSRYSGFDNLELRAAFAQGVYFTLSQLFMQT 497
QY 522 KHVATNTEVGNKHLNKNRSNNIELALGYEGDRWQYNLALYRNFPGNYYIAQTLDNRGP 581
D 498 SAGSVTY--GNPDLKAEHSNNFELGARYNGNWLIDSAVYYSKADYI--ASLICDGSIV 554
QY 582 KSIEDDESEMCLVRY--NQSADGFYCAEGEYFKPTPRYRIGVSGDYVRGRL----- 630
D 555 CNGNTNSRSRYYYDNDIDRAKTWGLE-----ISAENYNGWVFPYISGNLIRQYETST 608
QY 631 -----KNLPSLPGR-----EDAYGNRPPIAODDNA--PRVPA-ARLGFH 667
D 609 LKTTNTGEPAINGRIGLKHILVMQANILSDVFIARAASSAKDSDNGTETNPGWATLNFA 668
QY 668 LKASLTDRIDANLDYRV-PAQNKLA--RYET 696
D 669 VNTEF-----GNEDQYRINLALNLTDKRYRT 695

RESULT 17
Q8X8U8 PRELIMINARY; PRT; 719 AA.
AC
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative outer membrane receptor for iron compound or
DE colicin.
GN Z3159 OR ECS2792.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and
RT genome comparison with a laboratory strain K-12.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
DR EMBL; AE005425; AAG57058.1; -.
DR EMBL; AP002559; BAB36215.1; ALT_INIT.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 719 AA; 79681 MW; 39B5EFDAF10E46C0 CRC64;
```

RC	STRAIN=Pm70;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
----	--------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

RC	STRAIN=Pm70;				
RX	MEDLINE=21145866; PubMed=11248100;				
RA	May B.J., Zhang Q., Li L.I., Paustian M.L., Whittam T.S., Kapur V.;				
RT	*Complete genomic sequence of Pasteurella multocida Pm70.*;				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).				
DR	EMBL; AE006093; AAK02660.1; -				
DR	InterPro; IPR000531; TonB_boxC.				
DR	Pfam; PF00593; TonB_dep Rec; 1.				
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 742 AA; 84910 MW; 9B2814BDAAD229D1 CRC64;				
	Query Match				
	Best Local Similarity 20.8%; Score 253; DB 16; Length 742;				
	Matches 165; Conservative 117; Mismatches 291; Indels 222; Gaps 39;				
QY	67	RQKAVNLGDALDGVPGIHASQYGGASAPVIRGQTRRIKVLNHHGETGMDAFS-----	121		
Db	61	QQQADNADLNLILPGVNA-----GGFRP--GGQT-----LNING-MGDAEDRVQLDG	107		
QY	122	-----PDHAIWDTALSQQVEILRGPVTLIYSSGNVAGLVADGKIPKMPEN-GV	172		
Db	108	ATKSPKYOQSGSIFIPPELLRKVTVDKGNVSPQYGNFGAGTVKFTKDATDFLKENQKI	167		
QY	173	SGELGLRLSSGNLEKLTSGGINI-GLGKNFVLHTEGLYKSGDYAVPRYRNLRKLPDSPR	231		
Db	168	GGLFKYGNNSNNQKTYSTALVLQNEQKNIIDLLFGSVRNASNYTRP---DKSKILFSKN	224		
QY	232	RFANG-----QHRAVL-----GWRKFRYRTYSDRRDQVGLPAHSHYDDCH	273		
Db	225	NQKSLGIKVNWQITPEHLLTLFSVGIHKGW-----EPWAAKEDVMSRPTET-EIKRYG	277		
QY	274	ADIIQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHANGKPMIDLNRK	333		
Db	278	IDVAKRKLIVR-----DQKDESYSL-----KYRYPENNNKWINLSVQ	315		
QY	334	RYELRAEWKOPPGFEALRVHLNRNDYHDEKA-----GDAVENFENNQT	378		
Db	316	LSYSKTE-----QNDTRHEKVTSSFLGTIGNKSWITYSLDITDITST	359		
QY	379	QNARIELRHQPIGRKGS--WGVQYLGQKSSALSATSEAVK-----QPMLLDNKVQ	427		
Db	360	LN-----IGRAHEHLLFGLQWLKNKRNITLMYHKEGVKKADYNGYFQPYMPSGRQ	410		
QY	428	H-YSPFGVQAQWNDNTELEGGVRVEKQKASIRYD-----KALIDRENYKQPLDPLGA	479		
Db	411	YTHAFYLDQIKWQNFLEFTGG-----IRYDHINNIGQKNLAPRYN-----DISA	454		
QY	480	HRQTARSFALSGNWY---FTPOHKLSLTASHQE--RLPSTQELYA--HGKHVAINTFE	530		
Db	455	GHDYSQKNTNGWSYILGLKAYDVNHYLSLFTNFSKTWRAPVIDEQEYTOYSQASVSATSLN	514		
QY	531	VGNKHLNKRNSNIIEALGY---EGDRWQYNLALYRNRFNGYIYAQTLNDGRGPKSIEDD	587		
Db	515	LEKEMINQTRVGGI-ITLNLHFQENDAFQFRITTYFYNRGKNEIFKT-----RGVNCVENA	568		
QY	588	SEM-----KLVR--YNQSGADFYGAEGEYIFKPTPRYRIGVSGDYVRGLKNLPSLPGR	639		
Db	569	LVDVNSVCPIIISNVRNLPFGYVIOGALEAYTQSSYLFQ-GLTYSYVKGKEDTSRPNPWS	627		
QY	640	EDAYGNRPFFIAQDDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVVFAQNKLYARTRP	699		
Db	628	KTS-----TWIA---ETPPRKATATLGFNIPEY-----YFTAGWRAEFVRKQDRSP	670		
QY	700	-----GHHMLNLGANYRNTRYGEW-----NWYVKADNLLNQSVYAHSSFLSDTP	744		
Db	671	LSNDSKASYWALPSSKGYSLHSVFSWSPTKIKGMNFKVTVDNLNFRPY-----P	721		
QY	745	QNGRSFTG-GVNVKF	758		
Db	722	YLGELASGTGRNVKF	736		

RESULT 19
 Q9HYX3 PRELIMINARY; PRT; 721 AA.
 AC Q9HYX3; 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Probable TonB-dependent receptor.
 GN PA3268.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=ATCC 15692 / PA01;
 RC MEDLINE=20437337; PubMed=10984043;
 RX Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.L., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004749; AAC0656.1; -;
 DR InterPro; IPR00531; TonB_boxC.
 DR InterPro; IPR006195; TRNA_ligase_II.
 DR Pfam; PF00593; TonB_dep Rec; 1.
 DR PROSITE; PS00862; AA TRNA_LIGASE II; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 DR Receptor; Complete proteome.
 SQ SEQUENCE 721 AA; 79371 MW; 70E6716698A9E05D CRC64;
 Query Match 6.1%; Score 246; DB 16; Length 721;
 Best Local Similarity 21.2%; Pred. No. 6e-09;
 Matches 184; Conservative 102; Mismatches 277; Indels 304; Gaps 42;
 QY 4 TTLLKPIVLSILLINTPLLAQAHETEQSVGLTETVTVGKSRPRATSGLLTSTASDKII-S 62
 DB 5 SSLSRLAVALGVSMAQAQAEBAKELG--TWTVGDLGEADQAVVQNHGPGARSVVR 62
 QY 63 GDTLRQKAVNLGALDVGPHGTHASQYGGASAPV-----TRGTGRIKVLNHHGETGDM 117
 DB 63 REMLESQAQNRDVLKVPVGVQVQDNGTGGSDISLVNVRGGLTSR----- 108
 QY 118 ADFSPDHAIWVD-----TALS-----QQVEILRGFTLLYSSGNVAGLVD 157
 DB 109 --LSPRSTVMDGVPAAVAVPGQPLSMPLSIGNLESIDVVRGAGSVRGPNVGVGIN 166
 QY 158 VADGKIPKMPGVSGELGRL---SSGNLEKLT-----SGGINIGLGNFVLHTEGLYR 210
 DB 167 FVTRAIPEK-----SGEIGTTIEHAGHGKWKLNQAFGLGTADNGLG--VALLYSGV-- 217
 QY 211 KSGDYAVPRYRNKRLPDSPPRPANGHRAVLGWRKEFYRTYSDRDQYGLPAHSHEYD 270
 DB 218 KGADY-----RGNDNDIDVLLKTHWGLTDSQ-----LAANFHY-- 255
 QY 271 DCHADIWQSLINKRYLQYPLHLLTBEDVDYDNPGLSCGCFHDDDDAHAAHNGKWIIDL 330
 DB 256 DAVADM-----PGGLTQAYD--DDPFQSV-----RWDNDF 284
 QY 331 RNKR-----YELRAEWKQPFQFEALRVHLNR-----NDYHDEKAGDA 369
 DB 285 RGRKDFSLKYTRQVDDLTQFVLTYYSDSFRGSSIAARNLRTITSYPRDYHVAVEPRV 344
 QY 370 VENFFNQTQVARIELRHQPIGLRKGWSGVQYILGOKSALSATSSEAVKQWPLDN----- 424
 DB 345 SRIFAGPTT-----QEVG-----IGRYLKE-----ANNERASQALVDNPTVR 385
 QY 425 -----KVQHSYFFGVEQANWDFNFTLEGVVRVEKQKASIRYDKALI----- 464

RESULT 19
 Q9HYX3 PRELIMINARY; PRT; 721 AA.
 AC Q9HYX3; 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Probable TonB-dependent receptor.
 GN PA3268.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=ATCC 15692 / PA01;
 RC MEDLINE=20437337; PubMed=10984043;
 RX Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.L., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004749; AAC0656.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR InterPro; IPR006195; TRNA_ligase_II.
 DR Pfam; PF00593; TonB_dep Rec; 1.
 DR PROSITE; PS00862; AA TRNA_LIGASE II; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 DR Receptor; Complete proteome.
 SQ SEQUENCE 721 AA; 79371 MW; 70E6716698A9E05D CRC64;
 Query Match 6.1%; Score 246; DB 16; Length 721;
 Best Local Similarity 21.2%; Pred. No. 6e-09;
 Matches 184; Conservative 102; Mismatches 277; Indels 304; Gaps 42;
 QY 4 TTLLKPIVLSILLINTPLLAQAHETEQSVGLTETVTVGKSRPRATSGLLTSTASDKII-S 62
 DB 5 SSLSRLAVALTALGVSMQAQAEEAEKELG-TVTVGDMGLGEADQAVVQNHGPGARSVVR 62
 QY 63 GDTLRQKAVNLGALDVGPHGTHASQYGGASAPV-----TRGTGRIKVLNHHGETGDM 117
 DB 63 REMLESQAQNRDVLKVPVGVQVQDNGTGGSDISLVNVRGGLTSSR----- 108
 QY 118 ADFSPDHAIWVD-----TALS-----QVEILRGFTLLYSSNGVAGLVD 157
 DB 109 -LSRSTVMDIGVPAAVAPVGPQLSMPLSIGNLESIDVVRGAGSVRGPNVGVGIN 166
 QY 158 VADGKIPKMPGVSGELGRL-----SSGNLEKLT-----SGGINIGLGNFVLHTEGLYR 210
 DB 167 FVTRAIPEK-----SGEIGTTIEHAGHGWMKLNQAFGLGTADNGLG--VALLYSGV-- 217
 QY 211 KSGDYAVPRYRNKRLPDSPPRPANGHRAVLGWRKEFYRTYSDRDQYGLPAHSHEYD 270
 DB 218 KGADY-----RGNDNDIDVLLKTHWGLTDSQ-----LAANFHY-- 255
 QY 271 DCHADIWQSLINKRYLQYPLHLLTBEDVDYDNPGLSCGCFHDDDDAHAAHNGKWDIL 330
 DB 256 DAVADM-----PGGLTQAYD-DDPFSQV-----RWDNF 284
 QY 331 RNKR-----YELRAEWKQPPGFEALRVHLNR-----NDYHDEKAGDA 369
 DB 285 RGRKDFSLKYTRQVDDLTQFEVLTYTSDSFRGSSIAARNLRTITSYPRDYHVAVEPRV 344
 QY 370 VENFFNQTQNAITELRHQPIGLRKGWSGVQYILGOKSALSATSSEAVKQPMILDN----- 424
 DB 345 SRIFAGPTT-----QEVG-----IGRYLKE-----ANNERASQALVDNPTVR 385
 QY 425 -----KVQHSYFFGVEQANWDFNFTLEGVVRVEKQKASIRYDKALI----- 464

QY 221 --RNLKRLPSPRRFANGQRAVLGWRKRFYRRYTSYDRRQYGLPAHSHYDDCHADIW 278
D 222 DANNL-----RLNGEH-----LTDSAYRNQGLAKFGWQADDA-----265
QY 279 QKSLINKRYLYQLYPH-----LLTEEDVDYDNPGLSCGPHDDDDAHAAHNG 324
D 266 QRLSLVRQETHQNAFNSQWSTVPLVQOQTRDF-NTLLRYLNPTDSR-----317
QY 325 KPWIDLNKRYELRAEWKQPPGPEALRVHLNRNDYHHDEKAGDAVEN-----372
D 318 --WLDARVAVNKT-----FDEYRVTKQSDKVDYRTGLNLNHSQFDLLSLTYG 368
QY 373 --FNNQTONARIELRHOPI---GRLLGSGWQVQLGKSSALSATSEAVQPMLLDNKVQ 427
D 369 GDYIEDTKOGER-EGKRPPIPADGRSK-VWG-----SYVQADIPLGSOVLLPG---415
QY 428 HYSFFGVEQANWNTLEGGVVEKQKASIRYDKALIDRENYKQPLDPLGAHQRTARSF 487
D 416 -----LRYDHTAE-----DKNIAGSERSEHLSFVSGL-----444
QY 488 ALSGNWFTPOHKLSTASHQE---RLPSTOELYAHGKH-----VATNTFEVGNKHLNKR 540
D 445 ---RWAATDW--LTLNARYDEAFRAFSMEEMTYTGFHFCMPGMCNVFK-ANPDLKPE 497
QY 541 SNNIELALGYE-----GDRQVYNALYRNPFNGYIYAO-----TLNDGRGPKSIE 585
D 498 AKNKEISAKQRFNVLADDELAFATYFHNNSVNIIDQVNMFTTQVYVNTDAR-LRGVE 556
QY 586 DSEMKLVRYN-----QSGADPYGAEGEYFKPTPRYRIGVSGDYVRGLKNLP--SLPGR 639
D 557 LDA-----RYNWRDLETSLSAQTEG-----RDKTNPQINNIPAH 592
QY 640 EDAYGNRPFTAQDONAPRPAARLGFHLKASLTDRIDANLDYRVPAQNKLARYETRT- 698
D 593 KWLGVSHYFMDRD-----LKAGV-----NVSHVE--SQDELPSNVTVY 630
QY 699 PGHMLNGANRYNRTRYGEWNWVKA-----DNLNQSVAHSSFLSDTPQWGRSFT 751
D 631 PSYTLVDLYTWQPT-----GALKAVKVDVGDINDVIDE---YRQAP-DQLYSAGRNF 681
QY 752 GGVNVKF 758
D 682 GGVRYSF 688

RESULT 21
Q8EA32 PRELIMINARY; PRT; 708 AA.
AC Q8EA32;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TonB-dependent receptor, putative.
GN S04077.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Sehadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis";

RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015838; AAN57051.1; -
DR TIGR; S04077; -
KW Receptor; Complete proteome.
SQ SEQUENCE 708 AA; 78222 MW; 20E558517FFAIDF CRC64;

Query Match 6.0%; Score 244; DB 16; Length 708;
Best Local Similarity 20.4%; Pred. No. 8.1e-09;
Matches 171; Conservative 116; Mismatches 312; Indels 240; Gaps 36;

QY 9 IVLSILLINPLLAQAHEHETOSVGLFTVTVVVKSRPRATSGLLHTSTASDKIISGDLRQ 68
D 21 LTLALGLSPAP-LAMANDNEQT-QIERILVHGE-----QSICR 56
QY 69 KAVNLGDALDQVPGIHAQSYGGASAPVIRGQTRRIKVLNHHGETGMADFSFDHAIMV 128
D 57 NALGSADALLKQDQVDFSEAGVSALPVLNGMGDRILKVL-----VDGADITASCANQM 110
QY 129 DTALS-----QOVELRGPVTLILYSSGNVAGLVVDVADGKIPEKMPENGVSGBELGURLS 181
D 111 NPPLSYVSANQIINSVEVAVGSPVSAGGDIAGIKV-NSLNPKFTDSENLSFESG-EIS 168
QY 182 SGNLEKLTSGGINIG-----LGKNFVLHTEGLYRKSGDYAVPRYNLKLKLPDPSRRPANG 236
D 169 SG--YRSTSDSLIGAKAGIASKNVLSYOGAVEDANSY---HDGNGDKVLDITLYRAQN- 222
QY 237 QHRAVLGWRKRFYRRYTSYDRRQ-----YGLPAHSHYDDCHADIWQKSLIN 284
D 223 -HALTAAR-----DEKQQLAVLTHQAIPFGFPNQYMDMTD-----N 260
QY 285 KRYLOLYPHLLTREDVDYDNPGLSCGPHDDDDAHAAH-----NGKPWIDLRNKRY 335
D 261 KSYGALVRYLR-----DLENDG---EFSQALNWHSVKHEMGFFTPPEKTGKMPNTEGSDY 312
QY 336 ELRAEMKQPPGPEALRVHLNRNDYHDE-----KAGDAVENFNNQTONARIELR 386
D 313 SYQLHRLTWGDDSTLLGQEYYSYQLDDIWPVPGTMAPNDYININDGERRAAV--- 369
QY 387 HQPIGRKLSGWQVQLGKSSALSATSEAVKQPMLLDN--KVQHSYFFGVEQANWNTFL 444
D 370 -----YGEW-----QQNLNPLMWLSAGVRYEYVTTDTGCVQAYS-----NWPM 407
QY 445 EG--GVYVEKQKASIRYDKA-----LIDRENYKQPLDPLGAHQRTARSFALSGNWTTPQ 498
D 408 MGMPNVDAEAAKAFNAMDRSRDNLIDA-----TLIARYQLSAK 446
QY 499 HKLSLTASHOERLPSTOELYAHGKHVATNPF-----EVGNKHLNKRSSNNIELALG 549
D 447 QQLFGLARKNRPALNLYERSWGRGVWATTWIGWYGDGNGYVGNPDLKPETAHTLSAAYK 506
QY 550 YEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKSIEDDSEMKLVRYNQSGADFYGABGEI 609
D 507 FNGDAWQFSTAWYSAVTDYDAEVI--GSFNRTSTPAGKRNILKFTNEDANLYGAKFSA 564
QY 610 YEKETPRYRIGVSGDYVRGLKNLP--L-----PQREDAYGNRPFAQ-----DD 653
D 565 LY-----LLADTDSGKQWMLKLNITRGERDEGNEPLYQIKPLQTLALSHOLDGW 615
QY 654 QNAPRVPAARLGFHLKASLTDRIDANLDYRVPAQNKLARYETRTPGHMLNLCANRYRN 713
D 616 EN-----RLAQWVAT-KDRVDDR-----RLENETSSYSLNLSS-----S 650
QY 714 TRYGEWNWYKADNLLNQSVYA-----HSSFLSDTPQWGRSFTQWGVNVKF 758
D 651 IKMQELSLTAPAINNLF-DTYELPLGGSVIAEFKADSSNGFSQVAGSGRSFELGASVRF 708

RESULT 22
O66812 PRELIMINARY; PRT; 680 AA.
ID O66812
AC O66812;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

SQ	SEQUENCE	756 AA; 83202 MW; B499B739D04E25D5 CRC64;
DT	Query Match	5.8%; Score 235.5; DB 2; Length 756;
GN	Best Local Similarity	20.9%; Pred. No. 3.7e-08;
DE	Matches 176; Conservative 119; Mismatches 38; Indels 213; Gaps 37;	
QY	20	LLAQAHETEQQSGLETVTVVVGKSRPRATSGLLHTSTAGDKIISGDTLRQKAVNLGDALDG 79
DB	15	IVLEKRSAGSALAQQVTSIATRNEQDYNVPSTVSVHDR---AELDRQHVNNIRELVRY 71
QY	80	VPGIHASQYG--GGASAPVIRGQTGRRIRKVLNHHGETGD---MADFSPDHAIIMVDLTALSQ 134
DB	72	EPGVSVGAGTRSGNAGYINRIGIDGRVLTQDVGVEPDFNFNGFYAKTRRNYVDPEIVK 131
QY	135	QVEILLRGPVTLIYSSGNAVGLVD--VADGKPEKM--PENGVSGBELGLRLSGNLEKLT 190
DB	132	RVEILLRGPASALYSSAGNAVSYFTLD--PDDIIFKPGQDVGARLKTGYSSADESWLT 188
QY	191	GGINIGLGNF--VLHTEGLYRKSG---DYAVPRYRNLKELPSPRRFANGQHRVILGW 244
DB	189	GTF-AGRVQDPGLLH---LSQRNGHETESYDGNATGLARTGANPED-ARTNVLAKLW 243
QY	245	KRFYRTYSRRQYGLPAHSHYEDDCHADIIWOKSLINKRYLQLYPHLLTEEDVDYDN 304
DB	244	---NYG---DDNRLGLITVEK----- 257
QY	305	PGLSCGFHDDDAHAHANGKPMI-----DLRKNRYELRAEWKQPPPGFE 349
DB	258	-----FKDDRVLNKNVAGPFGGGRGNLYRRRRGNDTITRERFGLNTEFALESPIAD 311
QY	350	ALRVHLNRNDYHDE---KAGDAV-----ENFFNQTONA----- 381
DB	312	RIKTSLVNQIAKTDQTTAEIYQAGRVLRTRDTLYEEKQWFDQADLKAFSLGETDHQVT 371
QY	382	RIELRHQPI--GRUKSGWQVYLOKSSALSA---TSEAVKQPM-LLDNKKVQHYSPFGV 434
DB	372	YGTTLKQKQVTSREGSASCLAIGAAGCTAIGAPSPSASDVKKASDFDPDPTINTYSLFAQ 431
QY	435	EQANWDMFTLBGGVRVEKQASIRYDKALIDREN---YKQPLDPLGAHROTAR---SFA 488
DB	432	DOITWDKWTFPAPVRYDTRLKPQLTOEFLNVPNTGAYTVGDKDKTWNRVTPKFLTYA 491
QY	489	LSGNWYFTPQKLSLASHQBLRSTQELYAHGHVATNTFEVGNKHLNKNERSNNIELAL 548
DB	492	LTDNVTWFGQY-----AEGFRTPSAKALYGRFENLNGYTVENPDLKPTSKGIETGI 545
QY	549	GYEGRWQYNIALYRNRPGVI-----YAQTLNDGRGP-KSIEDDSEKMLVRY 595
DB	546	RGKFDEGSFDIAVYNYKYRDFIDEDKPVAGGTVEQFQAVNIKRATIKGVEAKGRNL--- 602
QY	596	NQSGADFYGABGEIYFKPTPRYRIGVSGDYVRGR-----LKNLPSLPGBEDAYGNR-P 647
DB	603	-----DTLGAPKGYTQGSVAYTYGRNDD--NGEFLNSVNLKGVFGYQDQDNYGVLVS 655
QY	648	PIAQDDQAPRVPAAARLGFHLKASLTDRIDANLDLYRFAQNKLARVETRTFGHHMLNG 707
DB	656	WTVVKQN--RVDSTT--FHAPDGGTDG-----PFKTPGFGILDLT 692
QY	708	ANYR-----RNRTRYGEWNW-YUKADNLLNQSVTAHSSFLSDTPQMGRSFTG 752
DB	693	AYYKVSQDVTVNGGLYNLTDKKY--WNWDDVRSYDSVGEAGVTGPANLDRLTQPGRNF- 748
QY	753	GVNV 756
DB	749	AINV 752

RESULT 26
Q9HV88
ID Q9HV88
AC Q9HV88
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
Probable outer membrane hemin receptor.	
PA4710.	
Pseudomonas aeruginosa.	
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
Pseudomonadaceae; Pseudomonas.	
NCBI_TaxID=287;	
[1]	
SEQUENCE FROM N.A.	
STRAIN=ATCC 15692 / PAO1;	
MEDLINE=20437337; PubMed=10984043;	
Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,	
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,	
Garber R.L., Coltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,	
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,	
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,	
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;	
"Complete genome sequence of Pseudomonas aeruginosa PAO1, an	
opportunistic pathogen."	
Nature 406:959-964(2000).	
EMBL; AE004885; AAG08096.1; -.	
InterPro; IPR000531; TonB_boxC.	
Pfam; PF00593; TonB_dep_Rec; 1.	
Receptor; Complete proteome.	
SEQUENCE 764 AA; 84724 MW; EA45897F2334255A CRC64;	
Query Match	5.7%; Score 231.5; DB 16; Length 764;
Best Local Similarity	19.8%; Pred. No. 7.3e-08;
Matches 172; Conservative 126; Mismatches 325; Indels 247; Gaps 35;	
QY	6 LKPIVLISLLIN-----TPLLAQAHETEQQSGLETVTVVVGKSRPRATSGLLHT 53
DB	9 LRP-CLALLLSPSLALAGNAVPLTPTTITRTEQAQVDSVPSTVQTRQLD----- 61
QY	54 STASDKIISGDTLRQKAVNLGDALDVGPIHASQVGGASAPV---IRGOTGRIRKVLN 109
DB	62 -----RQNVNNIKELVRYEFGV--SVGAGQAGITGYNTRIGDGNRLTQI 106
QY	110 HGETGD--MADFSPDHAIIMVDLTALSQQVILRGPVTLIYSSGNVAGLVD----- 157
DB	107 DGVELPNDFSGPVAQTHRNVDPIVKRVEILRGPASALYGSNAIGGAVSYFTLDPSDI 166
QY	158 VADCKIPEKMPENGVSGBELGLRLSSG---NLEKLT-----GGINIGLGNFV 202
DB	167 IKDKG-----DVGARLKAGYESASHSLTTSATVAGRADDDGLLHYGRQGE 214
QY	203 LHTEGLYRKSGDYAVPRYRNLKRLPDSPPRFANGQHRVILGW-----KRFYRR 251
DB	215 TESNGHGGTG-----LSRSEANPEDADSYLLKLGWNYAEGSRFGVLVFEKYSQ 265
QY	252 TYSRRDOYGLPAHSHYEDDCHADI-----IWQK---SLINKRYLQLYPHLLTEE 298
DB	266 VDTQKSAYGP-----YDKGKPAIPSPMLPGMYQWRKGNLTLTRERYGLEHFLDLSQ 320
QY	299 DVDYDNPCLSGFFHDDDAHAHANGKPMIDLRNKRVELRAEWKQPPFGPALRVHLNRN 358
DB	321 VADRIQSLNYQLAKTQATREFY-----YPTIRKVLRT 355
QY	359 DYHDEKA---GDAVENFFNNQTONA---RIELRHQPIGRKLGSGWQYVLGQKSSALSA 411
DB	356 DTTYKERLWVPDSQLDKSFAIGCTEHLISYGINLKHQKVTGWRSGTGNLDTGADSPDA 415
QY	412 TSEAVKQPMLLDNKKVQHYSPFGVEQANWDMFTLBGGVRVE-----KQKA 455
DB	416 LERSDFP---DPTVKTYALFAQDSISWNDWTFPLGRYDTRMEPHITDFLRTMKQSQ 472
QY	456 SIRYDKALIDRENYKQPLDPLGAHQRTARSFALSGWYFTPQKLSLTASHQRLPSTQ 515
DB	473 NTAVDES---DKKWHRVSPKFG---VTYDFAOHTYTWY--GQYAQGF-----RPTAK 516
QY	516 ELYAGKHVATNTFEVGNKHLNKNERSNNIELALGYEGDRWQYNLALYRNFENGYIAQTL 575
DB	517 ALYGRFENLQAGVHIEPNNPKPEKSSQSFETGLRGKFDGSGFVAVFYNKTRDFIDEDAL 576

```

QY 576 N-----DGRGPKSIEDSEMKLVRYNQSGADF-----YGAEGYIFKPTPRYRIGVSG 623
Db 577 NTDSTGNGQOTQSNRIERAVIK-----GVELKGRLELGAFAQOGLYTGQSVAYAYGRNK 632
QY 624 DVVRGR-LKNLSLP-----GREDAYGNRPFTAQDDQNAAPRVPAAELGFLHKLASLT---D 674
Db 633 D--NGEPINSNVPLTGVFLGVDYDEADGN-----YGLLSWTLVPRKD 672
QY 675 RIDANLDYRVFAQNKLARYETRTPOGHHMLNLCANVR-----ENTRYGEWN 720
Db 673 RVDDS-----THTPDGTASQKFTGFGVLDLSAYYRLSKDLTLNAGLYNLTDKKVWLWD 727
QY 721 WTVKADNLLNQSVYAHSSFLSTPQMGSRF 750
Db 728 DVRYGDSVGEASALAPAN-IDRLSQGRNF 756

RESULT 27
Q8ZG14 PRELIMINARY; PRT; 665 AA.
AC Q8ZG14;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative outer membrane protein (Putative outer membrane
iron/siderophore receptor).
GN YP01313 OR Y1872.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
[1]
[2]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Meule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527 (2001).
[2]
SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611 (2002).
DR EMBL: AJ414147; CAC90143.1; -
DR EMBL: AE013889; BAM86423.1; -
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep Rec; 1.
DR PROSITE: PS00430; TONB DEPENDENT REC 1; 1.
KW Hypothetical protein; Receptor; Complete proteome.
SQ SEQUENCE 665 AA; 73909 MW; F5ACA23F30B06871 CRC64;

Query Match 5.6%; Score 228, DB 16; Length 665;
Best Local Similarity 20.7%; Pred. No. 1e-07;
Matches 152; Conservative 116; Mismatches 280; Indels 188; Gaps 34;

QY 24 AHETQSGLVETVTVVVKSRPRATSLG---LHTSTASDKIIISGDTLRQKAV-NLGDALDG 79
Db 29 AEKTYATPTDITVY-----TAGFQRIQDSAASTSVTREQENKAYRDVTDALKD 81
QY 80 VPGIHASQYGGASAPV-IRGTGRRIRKVLNHHGTGMDAFSP-----DHAIWVDTA 131

```

```

Db 82 VEGVWIT--GGSTSDISIRGMAKYTLIL-VNGKRVDTTRTPNSDGGSGIQGMLPPLA 138
QY 132 LSQQVEILLRGPVTLTYSSGNVAGLVADVADGKIPEK-----MPENGVSSELGLRLS 181
Db 139 AIDREIVRVGPMSSLYGSDMGCVINITRKVGKEWHGTVRADATLQEDSKSGDI----- 193
QY 182 SGNLEKLSGGINIGLGRNVLHTEGLY-RKSGDYAVPRYRNLKRLPDSRPFANGQHRA 240
Db 194 -PQTWAYASGPLIDGL--LGLKVSGLLSHRSEDKIIDGY-----NQMRNRGTATF 241
QY 241 VLGRKRFYRRYRSDRDOYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDV 300
Db 242 TL-----TPDDNNEFFDIGHYVQDRNSTPQRTLALNG---TNSDT 279
QY 301 DYDNPGLSCGFHDDDAHAHANGKPIWIDLNRKRYELRAEWKQPPFGFEALRVHLNRNDY 360
Db 280 QYDR-----NNYAVTHNG--YYDFGNS-----TSYIQDET 308
QY 361 HHDEKAGDAVENFFNNTQNAEIRHOPIGRLKSGMGVQYLGOKSSALSATSEAVKQPM 420
Db 309 RNPQRKMSVDNIFN--TQTSFLDNHTLI--LGGQYRYEELYDKGNQLPSASDL----- 359
QY 421 LLDNKKVQHYSPFGVEQANW---DNFTLGGVREVEKASIRYDKALIDRENYKQPLPDL 477
Db 360 ---KKLTRWSMALFAEDBQWMTNDFTLGGIRMDQD-----ONY----- 395
QY 478 GAHQRTARSFALSNNWYFTTPQHKLSLTASHQRLPSTQ-----ELYAHGKHVATNTEF 530
Db 396 GTH-WTPR--LYGVWHLADQWTLKGGVGGYRSPDLRQATDDWGQSGGKG-GLPALI 450
QY 531 VGNKHLNKNERSNNIELALGYEG-DRWQYNLALYRNRFGNYI-----YAQTLD 577
Db 451 LGNSNLKPERSTISQIEGILMDQEGMNASVTLYTFDKDKITEVRNCDITTTTGTQGVFN 510
QY 578 GRGPKSIEDSEMKLVRYNQSGADFYGAEGYIFKPTPRYRIGVSGDYVRGLKNLPSLP 637
Db 511 GINYKFIKD-----RINVDKAMTRGAETATFADINQAWSLATNYTTFQSEQKS----- 558
QY 638 GREDAYGNRPFTAQDDQNAAPRVPAAELGFLHKLASLTDRIDANLDYRVFAQNKLARYE-- 695
Db 559 ---GAFAQPL---NQPKHMLNGLTNWKTTFEDATWIRAN---YRGKASEYLNRTSMG 608
QY 696 TRTPGHHMLNLCANVR 711
Db 609 SRTPSYTFVDLGANYQ 624

RESULT 28
P72121 PRELIMINARY; PRT; 723 AA.
AC P72121;
DT 01-FEB-1997 (TremBLrel. 02, Created)
DT 01-FEB-1997 (TremBLrel. 02, Last sequence update)
DE Outer membrane protein C precursor (Outer membrane protein OPRC).
GN OPRC OR PA3790.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
[2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=96349120; PubMed=8760927;
RA Yoneyama H., Nakae T.;
RT "Protein C (OprC) of the outer membrane of Pseudomonas aeruginosa is a
copper-regulated channel protein."
RL Microbiology 142:2137-2144 (1996).
[2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX STRAIN=20437337; PubMed=10984043;

```

[illegible]

```

QY 378 T-QNARIELRHO---PIGRKLSGWQVYLGOKSSAL-----SATSEAVKQPMLLNKVQ 427
Db 354 TLKYGYNVHQIKPOAFNLSQFKIE---DKDA TEEDKKKRENEKIAKAYRLNPTK 410
QY 428 HYSFFGVEQAN-WDNFTLEGGVVEKOKASIRYDKALIDRENYKQPLDGLGAHOTARS 486
Db 411 TDTGAYIEATHEIDGFTLTGGLAYDRPKVKTBDGK-----TVSS 449
QY 487 FALSGNW--YFTQHKLSLTASHO--ERLPSTOE-LYAHGKH-----VATNTEFEVGNKHLN 537
Db 450 SSLNPSFGVWQPREHWSFGSNHYSRSPRLDALQTHGKGIISTADGT-----K 501
QY 538 KERSNNIELALGYEGDRWQYNLALYR-----NRFNGNIYAQTLDG-----578
Db 502 AERARNTTEIGFNNDGTFAANGSVFROTIXDALANPQRHDSVAVREAVNAGIKONGYE 561
QY 579 -----RGPKSIEDSEMKLVRYNOSGADFYGAE-GEIY-----FKP 613
Db 562 LGASYRTGGLTAKVGVSRSPRFYDTHPKKLLSANPE---FGAQTGRTWTASLAVREFK- 616
QY 614 TPRYRIGVSGDYVRGLKNLPSLPGRDAGNRPFTAQDDONAPRVPAARLGFHLKASLT 673
Db 617 NPMLEIGWRGYY-----QKATGSILAAGQKDRDGKLENVVRQGFVN-----659
QY 674 DRIDANLDYRVEAQNKLARYETRTPGHMLNLGANYRRNTRYGEMWYVKADNLLNOSV 733
Db 660 -----DVFNWPKLCKDTLN-----VNLNVN-----NVPDKFY 687
QY 734 YAHSSFLSDT-PQMGSRFTGGVNVKF 758
Db 688 YPHSQRTWNTLPGVGRDVRLLGVNYKF 713

RESULT 30
Q51162
ID Q51162 PRELIMINARY; PRT; 720 AA.
AC Q51162;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Re-regulated protein B precursor.
GN FRPB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H44/76; PubMed=7558339;
RX MEDLINE=96009786; PubMed=7558339;
RA Fettersson A.M., Maas, van Massenaar, van der Ley, Tommassen;
RT "Molecular characterization of FrpB, the 70-kilodalton iron-regulated
RT outer membrane protein of Neisseria meningitidis.";
RL Infect. Immun. 63:4181-4184(1995)
CC 1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; X89755; CAA61902.1; -
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 720 AA; 7F9B540A12A3DB2D CRC64;

Query Match 5.3%; Score 214.5; DB 2; Length 720;
Best Local Similarity 21.6%; Pred. No. 1.1e-06;
Matches 186; Conservative 101; Mismatches 315; Indels 259; Gaps 45;

QY 16 INTP-----LAQHETQSVGLEFTVTVGKSRPRATSGLLHTSTASDKIIS 62
Db 1 MNTPLFRLSLLSLTLAAGFAHAENNAKVLDVTVTKGDRQ-----GSKIRTNIV- 50
QY 63 GDTLRQK-----AVNLGDLDGVGIHSAQVGGASAPVI-----RQGTERRIKV-----107

```

```

Db 51 --TLQQKDBSTATDMRELLKEPSI---DFGGNGTTSQFLTRMGONGSVDIKVDNAYS 105
QY 108 ---LNHGETGDMADFPDHAIMVDLTALSOQVEILRGPVTLTYSSGNVAGLVADVADKIP 164
Db 106 SQILYHQGR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIL 146
QY 165 EKMP-----NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVLHTEGLY-----RKSG 213
Db 147 AKTVDAQDLLKGLDKWGVRLNSGFSASNEGVSGASVFGKEGNF-----DGLFSYNRNDEK 202
QY 214 DYAVPR-YRNL---KRLPSQ-----RRFANGQHRVILGWKRKYR-----251
Db 203 DYAGKGFNRVNGKTVPSIALDKRSYLAKITGTFGDDHRIVLSHMKDOHRIQVREE 262
QY 252 -TVSDRRDQVGLPAHSHEDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCG 310
Db 263 FTVGDKSSRINIDRQAPAYRE-----TTQSNINLAYTGKVLG 299
QY 311 FHDDDDAHAAHAKGFWIDLNRKY---ELRAEWKQFPQGFPEALRVHLNRNDVHHDEKAG 367
Db 300 FVEKLDANAYV-----LEKERYSADDSGTGYAGNVKGNPHNTRITTRGANFNDFSRLA 351
QY 368 DAVENFNQOT-QNARIELRHO---PIGRKLSGWQVYLGOKSSALSATSEAVKQPM---420
Db 352 E-----QTLKYGYNVHQIKPOAFNLSKESIPTEEKN-----CQKVDKPMEQQ 397
QY 421 -----LLDNKVQHYVSFFGVEQA--NWDNFTLEGGVVEKOKASIRYDKALIDR 466
Db 398 MKDRADEDTVHAYKLSNPTKTDGTGVYVEAIHIDGDTLTGLYDRFKVK--THDGKTVSS 456
QY 467 ENYKQPLDGLGAHQRTARSFALSQWNYFTPOHKLSTASHQERLSTQELYAHGKH---523
Db 457 SNLN-----PSFGVIWQPH-----HWSFSASHNY---ASRSPRL--YDALQTHGKGI 501
QY 524 -VATNTEFEVGNKHLNKERSNNIELALGYEGDRWQYNLALYRNFNGYIYAQTLDNGRGP 582
Db 502 SIADGT-----KAERARNTTEIGFNNDGTFAAN-----GSYFW-QTIKDALANP 544
QY 583 SIEDDSEMKLVRYNOSGADFYGAE-GEIYKPTPRYRIGVSGDYVRGLKNLPSLPGRD 641
Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASVYRTGGLTAKGVSV-----HSKPRFYD 593
QY 642 AVGNRPFIADDDQNAAPRVPAARLGFHLKASLTDRI-DANLDY-YRVEAQNKLARYETRTP 699
Db 594 THKDKLLSANPEF-----GAQVGRWTASLAYRFQNPNIQGR-----GRYVQKAT 640
QY 700 GHMLNLGANYRR-----NTRYGEW-----NWYKADNLLNOSVYAHSS 738
Db 641 G-SILAAAGQKDRKGNLENVVRKGFVNDVFANWKPLGKDTLNVNLNVNFKFYYPHQS 699
QY 739 FLSDT-PQMGSRFTGGVNVKF 758
Db 700 RWTNTLPGVGRDVRLLGVNYKF 720

RESULT 31
Q9JWB8
ID Q9JWB8 PRELIMINARY; PRT; 714 AA.
AC Q9JWB8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative iron-regulated outer membrane protein.
GN FETA OR NMA0453.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

```

RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA	Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA	Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA	Whitehead S., Spratt B.G., Barrell B.G.;
RT	"Complete DNA sequence of a serogroup A strain of Neisseria
RT	meningitidis Z2491.";
RL	Nature 404:502-506(2000).
DR	EMBL; AL162753; CAB83750.1; -.
DR	InterPro; IPR000531; TonB boxC.
DR	Pfam; PF00593; TonB dep Rec; 1.
DR	PROSITE; PS01156; TonB_DEPENDENT_REC_2; 1.
KW	Complete proteome.
SQ	SEQUENCE 714 AA; 78885 MW; A3BBC1C1IED5527 CRC64;
	Query Match 5.3%; Score 213.5; DB 16; Length 714;
	Best Local Similarity 22.0%; Pred. No. 1.3e-06;
	Matches 188; Conservative 103; Mismatches 308; Indels 257; Gaps 47
Qy	11 LSILLINTPL-LAQAHETEVSGLTETTVVGKSRPRATSGLLHTSTASDKIISGDTLRQK 69
Db	LSL::SLTLTAAGFAHAENANVALDITVKGDRQ-----GSKIRTNIV---TLQOK 55
Qy	70 ---A/NLGDALDGVPGIHASOYGGCASAPVI-----PGTGRRIKV-----LNHH 111
Db	56 DESTATDMRELLKEEPSI---DFGGNGTGSOFLTRMGONGSVDIKVDNAYSDSQILYHQ 112
Qy	112 GETGDMADFSPDHAIWDVTALSSQVEIIRGPVTLTYSSGNAGLVADVADGKIPEKMPE-- 169
Db	113 GR-----FIVDPALKVVSVQKG-----AGSASAGIGATNGAIATKDVAQ 153
Qy	170 ---NGVSGELGRUSSG-NLEKLTSGGINTGLGNFVLHTEGLYRKS----GDYAVPR- 219
Db	154 DLLXGLDRNWGVRLNSGFASNEGVSYGASFVGEKNF---DGLPSYRNDEKDYEAGKG 209
Qy	220 YRNL----KRLPDSP-----REFANGQHRAVLGWKRFRRYSYSDRDQY-GLP 263
Db	210 FRNNFGKGTVPYSALDKRSYLAKITGTFGGDDHRIVL-----SHMKOOHGRI 258
Qy	264 AHSHEYDCHADIITQKSLIN-KRYLIQLYPHLITTEEDVDYNPGLSCGFHDDDDAHAAH 322
Db	259 TVREEF-----TVGDKKERISMQRQAQSYRE-TTOSNTNLAYTKDLGFVEKLDANAIV- 311
Qy	323 NGKPWIDLNRKY---ELRAEWKQFPFGFEARVLHNLNRNDYHDHKAGDAVENFENNOT- 378
Db	312 -----LEKKRYSADDSCSGYGAGNVKGFPHNQITTRGANFNFDSELAEE-----QTL 356
Qy	379 QNARIERHQ---PIGRLKSGWGVOYLQOKSSAL-----SATSEAVQPMLLDNKVOHY 429
Db	357 LKYGINTVRHQIEKQAFINSQFKIE---DXEKATDEEKNKRENPKIAKAYELTMTPTKD 413
Qy	430 SFFGVEQAN-WDNFTLEGGVVERKOKASIRYDKALIDRENYVKQLPLDGAHQRTARSFA 488
Db	414 AGAYIEAHEIDGFLTLCGLYDRFKVKTHDK-----TVSSSS 452
Qy	489 LSGNW--YFTPOHKLSLTASHO--ERLPSTOE-LYAHGKH-----VATNTEFVGKHLNKE 539
Db	453 LNPFGVITWQPEHWWSFSAASHYASRSPRLYDALQTHCKRGIIISTADGT-----KAE 504
Qy	540 RSNNIELALGEGDRWQYNLALYRNFONYLYAQTLDNG-RGPKSIEDSEMVK----- 591
Db	505 RARTEIGFNTDGTFAAN-----GSYFW-QTIKDALANPQNRHDSVAVREAENVAGY 555
Qy	592 -----LVRYNSGADFYGAEIGEIFYKPTPRYRIGVSGDYVRG---R 629
Db	556 IKNHGYELGASYRTCGLTAKVGVSHSKRPFYDTHDKLLSANPFCAQVGTWTASLAYR 615
Qy	630 LKNIPLSP-----GR--EDAYGNRPFIADODQNAPRVPAARLGFHLKASLTDRIANLDYY 683
Db	616 FKN-PNLEIGWRGVYQKAVSGSILVAGQKDRSGKLENVVRQGFGVN----- 660
Qy	684 RVFAQNKLARVETRTPGHHMLNLGANRNRTRYGEWNVYKADNLLNOSVYAHSSFSDT 743

Db	661	DVFANWKLGLKDTLN-----VNLSVN-----NVFNTEFFPHSQRWNTWT	69
Qy	744	-POMGRSFTGGVNVKF 758	
Db	699	LPVGROVRLGVNVKF 714	
RESULT 32			
ID	Q47232	PRELIMINARY; PRT; 673 AA.	
AC	Q47232;		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)	
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	FYUA precursor.		
GN	FYUA.		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Escherichia.		
OX	NCBI_TaxID=562;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Escherichia coli Phi;		
RA	Rakin A., Saken B., Harmsen D., Heesemann J.;		
RT	"The pesticin receptor of Yersinia enterocolitica: a novel virulence		
RT	factor with dual function.";		
RL	Mol. Microbiol. 13:0-0(1994).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Escherichia coli Phi;		
EX	MEDLINE=95247659; PubMed=7730256;		
RA	Rakin A.V., Urbitsch P., Heesemann J.;		
RT	"Evidence for two evolutionary lineages of highly pathogenic Yersinia		
RT	species.";		
RL	J. Bacteriol. 177:2292-2298(1995).		
CC	-!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).		
DR	EMBL; Z35106; CAA84489.1; ..		
DR	HSSP; P05825; 1PEP.		
DR	InterPro; IPR000531; TonB_boxC.		
DR	Pfam; PF00593; TonB_dep.Rec; 1.		
KW	Membrane; Outer membrane; Receptor; Signal; TonB box.		
FT	SIGNAL 1 22 POTENTIAL.		
FT	CHAIN 23 673 FYUA.		
SQ	SEQUENCE 673 AA; 73665 MW; 218360F62C65C71F CRC64;		
Query Match	5.3%; Score 213; DB 2; Length 673;		
Best Local Similarity	21.8%; Pred. NO. 1.3e-06;		
Matches 181; Conservative 93; Mismatches 323; Indels 232; Gaps 42			
Qy	1	MAQTLKPIVLISLLINTELLAQAHETQSGLFTVTVGKSRPATSGLLHTSTASDKI 60	
Db	1	MKWTRLPLAUGGLL--PANAQTSQD---ESTLVVTASKQSRSA--SANNVSTV 53	
Qy	61	ISGDTLRQKAVNLGDALGV-PGIHASVGGGASAPV-IRGTGRRIKVLNHHGTGDMA 118	
Db	54	VSAPELSDAGVTASDKLPVLPLGLNIENSGNWLFTSLRGVSSAQ----- 99	
Qy	119	DF-SPDHAIWVD-----TALS--QQVEILRGPVTLIYSSGNVAGLVADVAGKIPE 165	
Db	100	DFYNPAVTLIYDVGVPQLSTNTIQTALTDVQSVELLRGPQGTLYGRSAQGIIINIVTQQ-PD 158	
Qy	166	KMP----ENGVSGEIGLRLSSGNLEKLTSGGINTEGLKFNVLHTSLYRK--SGDYAVPR 219	
Db	159	STPRGYIEGGVSSRDSYR-SKFNI-----SGFIQDGL-----LYGSVTLRLQVDDGDMINPA 209	
Qy	220	YRN-----LKRLPD-----SPRFANGQHRVAVLGRKGRFYRFTYSDR 256	
Db	210	TGSDDLGGTRASIGNVKRLRAPDQPMWGMFAASRECTRAQTQDAVGVN----- 258	
Qy	257	RQYGLPAHSHEYDDCHADIWQSLINKRVLQIYPHLLTBEVDVYDNPGLSCGFHDDDD 316	
Db	259	-----DIKGRKLSIDSGSPDPMRRT-----DSQITLSGKYTTDD- 293	


```
QY 317 AHAAHNGKPIDLNRKRYELRAEMKQ-----PFP-GFEALRVHLNRNDYHHDEKA---G 367
Db 294 -----WV-----FNLISAWQQHYSTFFSGSLIVNIPQRWNOVDQELRAATLG 337
QY 368 DA--VENFNNTQNAIRIELRHQPIGRKLGSGVGYLGQSSALSATSSEAVKQPMILLDNK 425
Db 338 DARTVDMVFLYRQNT-----EKLNSAYDMPYLSSTGYTTAETLAA----- 382
QY 426 VOHYSPFQVEQANW---DNFTLEGGVRVEKQKASTRYDKALIDRENYKQPLDILGAHQK 482
Db 383 ---YS---DLTWLTDTRFDITGGVRFHSHDKSSTQYHGSMLG-----NPFQDQKSD 428
QY 483 TARSFALSGNWFTTPQHKLSLASHQERLPSQELYAHGKHVATNTFFVGKHLNKRNS 542
Db 429 DQVLGOLSGYMLTDDRWV-YTRVAGCYKPSGVNI-----VPTAGLDAKFPVAKSI 479
QY 543 NIELALGYEGDRWQVNLALYRNRFGNYIYAQTLDNGRGPKSIEDDSEMKLVRYNOSGADF 602
Db 480 NYELGTRYE---TADVTLQAATF---YTHTKDQWLYSGPVGMQTLG-----NAGKADA 526
QY 603 YGAEGEIYFKPTPRYRIGYSGDYVRGRLLKNLPSLPQREDAYGNRPFFIAODDONAPRVPAA 662
Db 527 TGVELEAKWRFPAGSWDINGNVIRSEFTN-----DSELYHGHR-----VPFVPRY 572
QY 663 RLGFHLKASLTDRIIDANLDYRVRFAQNKLYARTETRPCHHMLNLGANYRNRTRYGE---- 718
Db 573 GAGSSVNGVIDTRYGA-----LMPRLAVNLVGFHYFD-GDNQLRQCTYATLDS 620
QY 719 --W-----NWVKADNLNQ--SVYAHSSFLSDTPQMRGSRSTGGVNVK 757
Db 621 LQWQATERMISVYVDNLFRDRYRTYGVYMGSSAVAVNMGRVTGINTR 669
```

RESULT 33

```
Q9JRC5 PRELIMINARY; PRT; 703 AA.
AC Q9JRC5;
CT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Putative siderophore receptor (fomB-dependent receptor,
DE putative).
GN FHUA OR NMB0293.
OS Neisseria meningitidis, serogroup B.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20187481; PubMed=10722605;
RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,
RA Tinsley C.N.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish neisseria meningitidis from the closely related pathogen
RL neisseria gonorrhoeae."
RL Infect. Immun. 68:2082-2095 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=2017555; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cattone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science; 287:1809-1815 (2000).
```

```
DR EMBL; AJ391269; CAB72049.1; -.
DR EMBL; AF002386; AAF40744.1; -.
DR TIGR; NMB0293; -.
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 703 AA; 78860 MW; 392CD05997AAAF2FA CRC64;

Query Match          5.3%; Score 212.5; DB 16; Length 703;
Best Local Similarity 20.7%; Pred. No. 1.5e-06;
Matches 182; Conservative 117; Mismatches 260; Indels 321; Gaps 52;

QY 12 SILLINTPLLA---QAHETEQSVGLTETTVVCKSPRATSGLLH-----TSTASDKLIISG- 63
Db 12 TLIIASFVPAADTQDNGEHYTTATLPTVSWGQSDTSVLKGVINTYDEAAVTRNGQLIKET 71
QY 64 ----DTLR-QKAVNLG-----DALDGVPGIHASQYGGASAPVIRGOTGRRIRKVLNHHGE 113
Db 72 PQTITLAIQKNKNGTNDLSSILEGNAGIDA-----AYDMRGES-----TFLRGF 117
QY 114 TGDMADFSPDHA-----IMVDTALSOQVEILRGPVTLTYSSGNVAGLVADVADGKIPKMP 168
Db 118 QADASDIYRDGVRESGQVRRSTANIERVEILKGPSVLRYGRTNGGGINMVKYANFKQS 177
QY 169 EN--GVSG-----ELGLRLSS--GNLEKLTSG--GINIGLGNKPVLT 205
Db 178 RNIGAVYGSNWRSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNMVSPSITVKL 237
QY 206 EGLYKSGDYAVPRYRNKRLPDP-SPRRFANQHRVAVLGWRKRFYRRRTYSDDRDQVGLP- 263
Db 238 DNGLKWGTQYT---YDNVERTPDRSP-----TKSVY-----DRGLGY 272
QY 264 ----AHSHEYDDCHADIWQKSLINKRYLQLYPHLLTBEEDVDYDNPGLSCGFHDDDDA 319
Db 273 RMGFAHRNDF-----VKDKLQWV-----RSDLEY----- 296
QY 320 HAHNGKPMIDLNRKRYELRAEMKQFPFGFEALRVHLNRNDYH--DEKAGDAVENFP-- 374
Db 297 -AFNDK-W-----RAQWQ-----LAHRTAAQDFDHFYAGSENGNLIKRNAYAM 336
QY 375 ---NNQTONARIELR-HQPIGRKLG--SWGVOYLQK-----SSALSAT----- 412
Db 337 QQTDNKTLSSNLTLDNGDYTGIFENHLIVGMDYSREHRNPTLGFSSAFSASINPYDRASW 396
QY 413 -SEAVKQPMILDN--KVOHYSPF--GVEQANW--NFTLEGGVRVEKQKASIRYDKALIDR 466
Db 397 PASGRLOPILTONRHKADSYGIFVQNIIFSATPDLKFLVG-----RYDKYTFNS 445
QY 467 ENYKQPLDILGAHQK--TARSPA--LSGNWYFTPOHKLSTLASHQERLPSQELYAHGKH 523
Db 446 ENKL-----TGSSRQYSGHSPSPNIGAVNINPVH--TLYASYNK--GFAPYGGRGY 494
QY 524 VATNTFEVGNKHLNKRNNIELAL--GYEGDRWQVNLALYRNRFGNYIYAQTLDNDRGP 581
Db 495 LSLIDTLSSAVFNADPEYTRQYETGVKSSMLDRLSTLSAY----- 535
QY 582 KSIEDDSEMKLVRYNOSGADFYGAEGEIYFKPTPR---YRIGVSGD----- 624
Db 536 -----QIERFN-----IRYRPPKNNPYIYAVSGKHSRSGVLSAIGOI 574
QY 625 ----YVGRLLKNLPSLPQREDAYGNRPFFIAODDONAPRVPAAARLGFHLKASLTDRI 679
Db 575 IPKKLYLRGSLGWQAK-----VVEDKENPDRV-----GIHL--NNTSNVTGN 615
QY 680 LDYRV-----FAQNKLYARTETRPCHHMLN--LGANYRNR-TRYGE 718
Db 616 L-FFRYTPTENLYGEIYGVTKGRYGNYSRKNKEVTLTGPARVDAMLGNHKNVNTFA- 673
QY 719 WNWYTKADNLNQSVYAHSSFLSDTPQMRGSRSTGGVNVK 758
Db 674 -----AANLILNQYKYSDSM-----PGNPRGYTARVNVRF 703
```

```

RESULT 34
Q9JXL3 PRELIMINARY; PRT; 714 AA.
AC Q9JXL3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Iron-regulated outer membrane protein FpB.
GN NMB1988.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Ufferback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002548; AAF42315.1; -.
DR TIGR; NMB1988; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Complete proteome.
SQ SEQUENCE 714 AA; 79150 MW; F8BAP5BDB14CDFE CRC64;

Query Match 5.3%; Score 212.5; DB 16; Length 714;
Best Local Similarity 21.5%; Pred. No. 1.5e-06;
Matches 187; Conservative 96; Mismatches 305; Indels 281; Gaps 47;

Qy 16 INTPL-----LAQAHETEQSGVLETVVVGKSRPRATSGLLHTSTASDKIIS 62
Db 1 MNTPLRLSLSLTLAAGFAHAENNAKVLDTVTVGDRQ-----GSKRTNV- 50
Qy 63 GDTLRQK-----AVNLGDALDGVPGIHASQYGGGASAPVI-----RGOTGRRIKV----- 107
Db 51 --TLQKDESTATDNRELLKEPSI---DFGGNGTSQFLTRGMQNGSVDIKVDNAYSD 105
Qy 108 ---LNHHGETGMADFPDPHAIMVDTALSOQVEILRGPVTLTLLYSGNVAGLVVDVADGKIP 164
Db 106 SQILYHQGR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIL 146
Qy 165 EKMP-----NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVILHTEGLY-----RKSG 213
Db 147 TKTVDAQLLGLDKNWGLRVNLSGFASNEGVSGASVFGKEGNF-----DGLFSYNRNNEK 202
Qy 214 DYAVPR-YRNL-----KRLPDSF-----RRFANGQHRVILGWKRFRVRYTSDRR 257
Db 203 DYACKGFRNPNFGKVTVPYSDLDKRSYLAIGTSFGDGRIVL-----SNMK 251
Qy 258 DQY-GLPAHSHEYDDCHADITWQKSLIN-KYLYQLYPLLFEEDVDYDNPGLSGCGFHDD 315
Db 252 DQHRGIRTVREEF-----TVGGDKERISMERQAPAYRE-TTQSNNTLAYTGKNLGFVEKL 305
Qy 316 DAHAHANGKWDILRNKRYE-----LRAEWQPPFGFEAL 351
Db 306 DANAYV-----LEKERYSADSGTGAGNVKPNHTQITTRGMNPNFDSRLAEQTL 357
Qy 352 RVHLNRNDYHDEKAGDAVENFFNNQTONARIELRHQPIGLKSGWGYQLGQKSSALSA 411
Db 358 KYGIN---YRHOEIKPQA---FLNSQ---FKIEDKEKATDEKKN-----NR 395
Qy 412 TSEAVKQPMLLDNKVQHYSPFFGVQAN--WDNFTLEGGVRVBEKQKASTRYDKALIDRENY 470

```

```

Db 396 ENEKIAKAYRLTNPTKTDGTGAYIEAIHEIDGFTLTGGLRYDRFKVK--THDGKTVSSNNLN 454
Qy 471 KOPLPDLGAHQRTARSFALSNGWTFPTPOHKLSLTASHOERLPSTOELYAHGKH-----VAT 526
Db 455 ---PSFGVIWQPHE-----HWSFSASHNY---ASRSPL--YDALQTHGKRGIIISAD 499
Qy 527 NTFEVGNKHLNKRSSNNIELALGYEGDROWYNLALYRNRFNGYIYAOTLNDG-RGPKSIE 585
Db 500 GT-----KAERARTEIGFNTYNDGTFAN-----GSYFW-QTIKDALANPQNRH 542
Qy 586 DDSEMK-----LVRYNOSGADFYGAEGEIFYFKPPTRYRI 619
Db 543 DSVAVREAVNAGYIKNHGYELGASVETGGLTAKVGVSHSKPRFYDTHDKLLSANPEGA 602
Qy 620 GVSQDYVRG-----RLKNLPSLP-----GR--EDAYGNRPPIAODDQNAVPVPAARLGFHLKA 670
Db 603 QVGRWTWTLASYRFQNP-PLNEIGWRGRYVQKAVGSILVAGQKDRNGKLENVVRKGFQGVN- 660
Qy 671 SLTDRIIDANLDYYRFAQNKLARVETRTPGHMLNLGANYRNTRYGEMWVYKADNLLN 730
Db 661 -----DVFAWKPLGKDTLN-----VNLSVN-----NVFN 685
Qy 731 QSVYAHSSFLSDT-PQMGSRSTGGVNVKF 758
Db 686 TFYYPHSQRWNTLPGVGRDVLGLVNYKF 714

RESULT 35
Q47231 PRELIMINARY; PRT; 673 AA.
AC Q47231;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Fyua precursor.
GN FYUA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K49;
RX MEDLINE=95075311; PubMed=79841105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RT factor with dual function."
RL Mol. Microbiol. 13:253-263 (1994).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K49;
RX MEDLINE=95247659; PubMed=7730256;
RA Rakin A.V., Urbitsch P., Heesemann J.;
RT "Evidence for two evolutionary lineages of highly pathogenic Yersinia
RT species."
RL J. Bacteriol. 177:2292-2298 (1995).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Z38065; CAA86212.1; -.
DR HSSP; P05825; 1FEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22
FT CHAIN 23 673
FT FYUA.
SQ SEQUENCE 673 AA; 73713 MW; 2635A53A7515BD08 CRC64;

Query Match 5.3%; Score 212; DB 2; Length 673;
Best Local Similarity 21.8%; Pred. No. 1.5e-06;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTTLPVLSILLINTPLIAQAHETEQSGVLETVVVGKSRPRATSGLLHTSTASDKI 60

```

```

Db 1 MKMTLYPLALGGLL--PTTAAQTSQD---ESTLVVTASKQSSRA--SANNVSTV 53
Qy 61 ISGDTLRQKAVNLGDGV-PGIHASQYGGASAPV-IRGTGRRRIKVLNHHGETGMA 118
Db 54 VSAPELSDAGVTASDKLPRVLPLGLNIENSGNMLFSTISLRGVSSAQ----- 99
Qy 119 DF-SPDHAIMVD-----TALS--QOVELRGPVTLIYSSGNVAGLVDAVGKPIE 165
Db 100 DFYNPAVTLYVDGVPQLSTNTIQAULTDQSVELLRGPQGTLYGKSAQGGIINIVTQQ-PD 158
Qy 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLKNFVHLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKPNL-----SGPIQDGL-----LYGSVTLRQVDDGDMINPA 209
Qy 220 YNN-----LKRLPD-----SPRRFANGQHRVGLGWRKRYRYRYSR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGFPAASRECTRAQDAVYGMN----- 258
Qy 257 RDQYGLPAHSHEYDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDD 316
Db 259 -----DVKRKLISDGSDDPYMRRCT-----DSQTLGKVTDD- 293
Qy 317 AHAHAHNGKPMIDLNRKRYELRAEWKQ-----PFFGFALRVLHNRNDYHHDEK----- 365
Db 294 -----WV-----FNLSAQOQHYSTFPF-GSLIVNMPQR-WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNNTQNAIRLHQPICRLKSGWGVQVGLGQKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYRQNT--DVKRKLISDGSDDPYMRRCT-----DSQTLGKVTDD- 293
Qy 424 NKVQHSFFGVQANW---DNFTLGGVVEKQKASIRYDKALIDRENYKQPLDGLGAH 480
Db 383 -----YS-----DLTWHLTDRFDIGGVRFHDKSSTQYHGMG-----NPPGDQGS 426
Qy 481 RQTARSAFALSGNWYPTPQKLSLTASHQERLPSTQELYAHGKHVATNTPEVGNKHLNKR 540
Db 571 RYCAGSSVNGVIDTRYGA-----LMPRLAVNLVGPYFD-GDNQLRQGTATLD 618
Qy 719 ---N-----NWYKADNLNQ--SVYAHSSFLSDTPQMGHSFTGGVNVK 757
Db 619 SSLGWAQTERMNISVYVDNMLFRRYRTYGYMNGSSAVAVNMGRTVGINTR 669

```

RESULT 36

```

Q57303
ID Q57303 PRELIMINARY; PRT; 673 AA.
AC Q57303;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE FYUA precursor.
GN FYUA
OS Yersinia pseudotuberculosis, and
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633, 562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YERSINIA PSEUDOTUBERCULOSIS IA;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;

```

```

Mol. Microbiol. 13:0-0 (1994).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=YERSINIA PSEUDOTUBERCULOSIS IA;
EX MEDLINE=95247659; PubMed=7730256;
RA Rakin A.V., Urbitsch P., Heesemann J.;
RT "Evidence for two evolutionary lineages of highly pathogenic Yersinia species.";
RL J. Bacteriol. 177:2292-2298 (1995).
CC -|- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Z35107; CA84490.1; -
DR EMBL; Z35105; CA84488.1; -
DR HSP; F05825; IFBP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 673 FYUA.
SQ SEQUENCE 673 AA; 73682 MW; 5BFF31DA7EA15DFE CRC64;

Query Match
Best Local Similarity 5.3%; Score 212; DB 2; Length 673;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTLKPIVILINIPILIAQAHEHTEQSGVLEFVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTLYPLALGGLL--PTTAAQTSQD---ESTLVVTASKQSSRA--SANNVSTV 53
Qy 61 ISGDTLRQKAVNLGDGV-PGIHASQYGGASAPV-IRGTGRRRIKVLNHHGETGMA 118
Db 54 VSAPELSDAGVTASDKLPRVLPLGLNIENSGNMLFSTISLRGVSSAQ----- 99
Qy 119 DF-SPDHAIMVD-----TALS--QOVELRGPVTLIYSSGNVAGLVDAVGKPIE 165
Db 100 DFYNPAVTLYVDGVPQLSTNTIQAULTDQSVELLRGPQGTLYGKSAQGGIINIVTQQ-PD 158
Qy 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLKNFVHLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKPNL-----SGPIQDGL-----LYGSVTLRQVDDGDMINPA 209
Qy 220 YNN-----LKRLPD-----SPRRFANGQHRVGLGWRKRYRYRYSR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGFPAASRECTRAQDAVYGMN----- 258
Qy 257 RDQYGLPAHSHEYDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDD 316
Db 259 -----DVKRKLISDGSDDPYMRRCT-----DSQTLGKVTDD- 293
Qy 317 AHAHAHNGKPMIDLNRKRYELRAEWKQ-----PFFGFALRVLHNRNDYHHDEK----- 365
Db 294 -----WV-----FNLSAQOQHYSTFPF-GSLIVNMPQR-WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNNTQNAIRLHQPICRLKSGWGVQVGLGQKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYRQNT--DVKRKLISDGSDDPYMRRCT-----DSQTLGKVTDD- 293
Qy 424 NKVQHSFFGVQANW---DNFTLGGVVEKQKASIRYDKALIDRENYKQPLDGLGAH 480
Db 383 -----YS-----DLTWHLTDRFDIGGVRFHDKSSTQYHGMG-----NPPGDQGS 426
Qy 481 RQTARSAFALSGNWYPTPQKLSLTASHQERLPSTQELYAHGKHVATNTPEVGNKHLNKR 540
Db 571 RYCAGSSVNGVIDTRYGA-----LMPRLAVNLVGPYFD-GDNQLRQGTATLD 618
Qy 719 ---N-----NWYKADNLNQ--SVYAHSSFLSDTPQMGHSFTGGVNVK 757
Db 619 SSLGWAQTERMNISVYVDNMLFRRYRTYGYMNGSSAVAVNMGRTVGINTR 669

```

```
Db 571 RYAGSSVNGVIDTRYGA-----LMPLRLAVNLVGHYFD-GDNQLRQGTATLD 618
Qy 719 ----W-----NWYKADNLLNQ--SVYAHSSFLSDTPQMGSRSTGGVNVK 757
Db 619 SSLGWQATERMNISSVYVNDLFDRRYRTYGYNGSSAVAQVNMGRVTGINTR 669

RESULT 37
Q47230
ID Q47230 PRELIMINARY; PRT; 673 AA.
AC Q47230;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Fyua precursor.
GN Fyua.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K235;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
factor with dual function.";
RL Mol. Microbiol. 13:253-263(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K235;
RX MEDLINE=95247659; PubMed=7730256;
RA Rakin A.V., Urbitsch P., Heesemann J.;
RT "Evidence for two evolutionary lineages of highly pathogenic Yersinia
species.";
RL J. Bacteriol. 177:2292-2298(1995).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Z38064; CA886211.1; -.
DR HSSP; P05825; IPEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 673 FYUA.
SQ SEQUENCE 673 AA; 73697 MW; C18FFB601EAL4110 CRC64;

Query Match
Best Local Similarity 22.1%; Pred. No. 1.5e-06;
Matches 185; Conservative 96; Mismatches 306; Indels 252; Gaps 46;

Qy 1 MAQTLKPIVLSILLINTPLLAQAHEHTEQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTLYPLALGGLL--PAIANAQTQSD--ESTLVVTASKQSSRSA--SANNVSVTV 53
Qy 61 ISGDTLRQKAVNLGDGV--PGIHASQVGGASAPV--IRGOTGRIRKVLNHHGTGDMW 118
Db 54 VSAPELSAGVATSDAKLPRVLPLGLNIENSGNMLFTSILRGSVSSAQ----- 99
Qy 119 DF-SPDHAIWVD-----TALS--QQVEILRGPVTLTYSSGNVAGLVADVADGKIPE 165
Db 100 DFYNPAVTLVYDGVLPQLSTNTIQTALTDVQSVELLRGPQGLTKSAQGGIINIVTQQ-PD 158
Qy 166 KMP-----ENGVSSELGLALSSGNLKLSTGGNIGLGRKRVLHTGLYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKENL-----SGPIQDGL-----LYGSVTLRLRQVDGDMINPA 209
Qy 220 YRN-----LKRLED-----SPRFANGQRAVLGWKRYRYTYSR 256
Db 210 TGSDDLGGTRASIGNVKURLAPDDQFWEFGFAASRECTRATODAYVGNW----- 258
Qy 257 RDQYGLPAHSYEDPDCHADIIMQKSLINKRYLQLYPHLLTBEDVDVDPNGLSCGFHDDDD 316
```

```
Db 259 -----DIKGRKLISDGSPPDPYMRRT-----DSQTLSGKYTTDD- 293
Qy 317 AHAHAHNGKPMIDLRNKYELRAEWKQ-----PPFGFEALRVHLNRNDYHHDEK----- 365
Db 294 -----WV-----FNLSAQQQHYSKRTFPS--GSLIVNMPQR--WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNNQTONARIELRHQPIGRFLKSGWGVQYLQKSSALSATSSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYRQNTNR-----EKLSAYDMPMTPLYLSSTGYTTAETLAA----- 382
Qy 424 NKVQHYSFPGVEQANW---DNFTLEGVVRVEKQASIRYDKALIDRENVYKQPLPOLGAH 480
Db 383 -----YS-----DLTWHLTDREFDGGVRFSHDKSKSTQYHGSMLG-----NPFQDGQKS 426
Qy 481 RQTARFALSGNWYFTPOHKLSTASHQERLPSTOELYAHGKHVATNTFEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDWRV--YTRVAQGYKPSGYNI-----VPTAGLDAKPFVAEK 477
Qy 541 SNNTELALGYEGDRWQYNLALYRNFGNYIAQTLDNGRGPKSIEDDSEMKLVRYNQSGA 600
Db 478 SINYELGTRYE---TADVTLOAATF--YTHKQWQLYSGPVGMQTLN-----NAGKA 524
Qy 601 DFYGAEGEYFEPKTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNR--PFIAQDDQNAAPRV 659
Db 525 DATGVLEAKRFPAGFQSWDINGVIRSEFTN-----DSELVHGNNRVPV-----PRY 572
Qy 660 PAARLGFHLKASLTDRIANDLYRVFAONKLARYETRTP-----GHHMLNLGANYRR 712
Db 573 GAG-----SSVNGVID-----TRYGALMPRLALNLVGPHYFD-GDNQLR 610
Qy 713 NTRYGE-----W-----NWYKADNLLNQ--SVYAHSSFLSDTPQMGSRSTGGVNVK 757
Db 611 QGTATLDSLGWQATERMNISSVYVNDLFDRRYRTYGYNGSSAVAQVNMGRVTGINTR 669

RESULT 38
Q8CVY3
ID Q8CVY3 PRELIMINARY; PRT; 673 AA.
AC Q8CVY3;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative pesticin receptor precursor.
GN C2436.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
KW Receptor; Complete proteome.
SQ SEQUENCE 673 AA; 73683 MW; 5BFF31DA7EA15DFE CRC64;

Query Match
Best Local Similarity 21.8%; Pred. No. 1.5e-06;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTLKPIVLSILLINTPLLAQAHEHTEQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTLYPLALGGLL--PAIANAQTQSD--ESTLVVTASKQSSRSA--SANNVSVTV 53
Qy 61 ISGDTLRQKAVNLGDGV--PGIHASQVGGASAPV--IRGOTGRIRKVLNHHGTGDMW 118
```

Db 54 VSAPSLDAGVTASDKLPRVLPLGLNIENSGNMLFSTISLRGVSSAQ----- 99
QY 119 DF-SPHAIMVD-----TALS-QQVEILRGVPTLLYSSGNVAGVVDVADGKPE 165
Db 100 DFYNPAVTVYDGVPLQSTNTIQALTDVQSVELLRGPGTLYGKSAQGGIINVTQQ-PD 158
QY 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLTGTEGLYRK--SGDYAVPR 219
Db 159 STPRGIEGVSSRDSYR-SKFLN-----SGPIQDGL-----LYGSVILLKQVDDGDMINPA 209
QY 220 YRN-----LKELPD-----SPRRFANGQRAVLGRKRRFYRRYTSR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPWEMGFASRECTRATQAYYGVN----- 258
QY 257 RDQYGLPAHSHEYDCHADIIWOKSLINKRYLQLYPHLLTEEDVDVNDPGLSCFFHDDDD 316
Db 259 -----DIKGRKLSISDGSPPPYMRCT-----DSQTLISKYTTDD-- 293
QY 317 AHAHAHNGKFWIDLRNKRVELRAEWKQ-----PFGPGEALRVHLNRNDVHHDEK-- 365
Db 294 -----WV-----FNLISAWQQOHSYRFFPS--GSLIVNMPOR--WQDVQOELRAAT 335
QY 366 AGDA--VENFFNMQTQARIELRHQPIGRKLGSGVQYLGQKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFGLYRQNT-----EKLNSAYDMPTMPLSSSTGYTTAETLAA----- 382
QY 424 NKVQHSYFPGVQANW---DNFTLEGGRVVEKOKASIRYDKALIDRENYKQPLDGLAH 480
Db 383 -----YS-----DLTWHLIDRFDIGGVRPSHDKSSQVHSGMLG-----NPFQDGKS 426
QY 481 ROTARFALSNGWYFTPOHKLSLTASHQERLPSTQELYAHGKHVATNTEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDMRV--YTRVAQGYKPSGYNI-----VPTAGLDKPFVAEK 477
QY 541 SNNIELALGYEGDRWQYNLALYRNFGNIYIAGTLNDGRGPKSIEDDSEMKLVRYNOSGA 600
Db 478 SINYELGTRYE---TADVTLQAATF--YTHTKDMQLYSGPVGQMQLS-----NAGKA 524
QY 601 DFYGAEGEYFKPTPRYRGSDYVGRKLNPLSLPGREDAYGNRPFFIAQDDQNAPRVP 660
Db 525 DATGVELEAKWRPAGWSWDINGNVIRSEFTN-----DSELYHGNR-----VFFVP 570
QY 661 AARLGFHLKASLTDRIANDLDYRVFAQNKLIARETTRTGHMLNLGANYRRNTRYGE-- 718
Db 571 RYGAGSSVNGVIDTRYGA-----LMPLRLVNLVGPYFD--GDQLRGQTYATLD 618
QY 719 -----W-----NWYVADMLLNO--SVYAHSSFLSDTPQGRSFTGGVNVK 757
Db 619 SSLGWQATERNMISVYVDNLFDYRRYRTYGYMNGSSAVAQYNNMGRVTGINTR 669

RESULT 39

Q9XD18 PRELIMINARY; PRT; 687 AA.
AC Q9XD18;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE Putative TonB-dependent outer membrane receptor protein.
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 9343;
RX MEDLINE=99307214; PubMed=10377135;
RA Comstock L.B., Coyne M.J., Tzianabos A.O., Pantosti A.,
RA Oerendonk A.B., Kasper D.L.;
RT "Analysis of a capsular polysaccharide biosynthesis locus of
RT Bacteroides fragilis";
RL Infect. Immun. 67:3525-3532 (1999).
CC -|- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).

DR EMBL: AF048749; AAD40727.1; --
DR InterPro: IPR005531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; 1.
KW Membrane; Outer membrane; Receptor; TonB_box.
SQ SEQUENCE 687 AA; 78838 MW; BBEB5E3502815E59 CRC64;
Query Match 5.2%; Score 211.5; DB 2; Length 687;
Best Local Similarity 22.1%; Pred. NO. 1.7e-06;
Matches 142; Conservative 84; Mismatches 220; Indels 197; Gaps 32;
QY 9 IVLSILLINT-PLLAQAHETEQSVGLETVTVGKSPRATSGLLTSTASDK----- 59
Db 6 LALLTLIGTLPLAAQNVREQ-----DVSVMND--PFNLQIVVTATREKIKNTPTVI 60
QY 60 --IISGDTLRQXAV-NLGDAL-DGVFGIHAHQYGGGASAPVIRGQTRRIKVL-----NHH 111
Db 61 TQIITSKQIEERGTGNIQDLTQEVPLGNFQEVGYGTSID-IQGLGSKHILFLIDGERIA 119
QY 112 GETGMDAESPDAIMVDTALSOQVEILRGPTVLLYSSGNVAGLVADVADGKIPEKM----- 167
Db 120 GENGGNIDYSRLNLYNID-----HIEIVKGASSALYGSQAMGGVINIITRKAKKFEASA 174
QY 168 -----PENGVSGBELGRLSSGNLEKLTSGGINIGLGNFVLTGTEGLYRKS 212
Db 175 GIRYAGNQOQYKDTPKDHSQYKRIHLDPNLTNLSGLNLG---KFTMTDVLKSF 231
QY 213 GDYAV-----PRY-----RNKRLPDSRRFANGQRAVLGWRKRFYRTYSDRR 257
Db 232 DGQYLFKPKPLVKYFPAYNTTITEELSKTPTISISGYEDVQVAHKMDY--RFSKRLKVQLK 289
QY 258 DOYGLPAHSHEYDCHADIIWOKSLINKRYLQLYPHLLTEED-----VDY---DNPGLS 308
Db 290 GSYM--LNVY--DFQADNIFEKS-----EDTYGGSIDYTTISDKSSLV 329
QY 309 CGFHDDDDAHAHNGKFWIDLRNKRVELRAEWKQFPFGPEALRVHLNRNDVHHDEKAGD 368
Db 330 ASVHTD-----HYNRYD-KVELKSGR 349
QY 369 AVENFFNMQTQARIELRHQPIGRKLGSGVQYLGQKSSALSATSE-AVKQPMLLDNKQV 427
Db 350 RLE--YKNNIIOPRIYVSTTALDKQITIGGLEI--YRESLFSDFKFTGVK-----ENKSQ 400
QY 428 HYSFFGVEQANWD---NFTLEGGRVE-----KQKASIRYD-KALIDRENYKQPL 474
Db 401 WTA-TAFLQDDWSINKQFSVIAGLRCDYHEKYGTNTLTPKASVMYKIFPFTVRFNY----- 454
QY 475 POLGAHQRTARSPALSGNWYFTPOHKLSLTASHQERLPSTQELYAHGKHVATNTEVGNK 534
Db 455 -----ARGY-----RSPSIKELYMNWDHLGM-FWIYVNS 482
QY 535 HLKERSNNIELALGYEGDRWQYNLALYRNFGNIYIAGTLND 577
Db 483 KLPETNNYISLSGEYVNSWINANVYNSWFRNKIEGMSND 525

RESULT 40

Q51132 PRELIMINARY; PRT; 697 AA.
ID Q51132
AC Q51132;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Iron-regulated outer membrane protein FtpB (fragment).
GN FtpB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2996;
RX MEDLINE=97124210; PubMed=8969523;
RA Der Ley P., der Biesen R., Sutmoller R., Hoogerhout P., Poolman J.T.;

RT "Sequence variability of FrpB, a major iron-regulated outer membrane
RT protein in the pathogenic neisseriae.";

Search completed: November 14, 2003, 11:00:06
Job time : 51 secs

Query Match 5.2%; Score 210.5; DB 2; Length 697;
Best Local Similarity 21.5%; Pred. No. 2.1e-06;
Matches 182; Conservative 88; Mismatches 314; Indels 261; Gaps 42;

Qy	26	ETBQSVGLTETVTVGKSRPRATSGLLHTSTASDKILISGDTLRQK----	AVNLGDAIDGVFP	81
Db	2	ENNAKVAIDTETVTKGDRQ-----GSKIRTNIV---TLQOKBESTATMDRELLKEEP	49	
Qy	82	GIHASQYGGGASAPVI-----RGQTRRIKV-----LNHHGETGDMADFPDHAJM	127	
Db	50	SI-----DFGGNGTSQFLIRGMGQNSVDIKVDNAYSQSILYHQGR-----	FI	94
Qy	128	VDTALSQQVEILRGPVUTLLYSNGNVAGLVDVADGKIPERKPE-----NGVSGELGLRASS	182	
Db	95	VDPALVKVVSQKG-----AGSASAGITGATNGAIIVKTVDGQDILLKGLDKNMGVRLNS	147	
Qy	183	G--NLEKLTSGGINIGLCKNFVLHTEGLY-----RKSGDYAVPR--VRNL-----KELPDSP--	230	
Db	148	GFASNEGVSYGASVFGKEGNP-----DGLFSYNRNDEKYEACKGFNFVTGGKTVPVYSALD	203	
Qy	231	-----RRFANGQHRAVILGWKRFXRYRTYSRRDQY--GLPASHSHEDDCHADIIWOK	280	
Db	204	KRSYLAKITGTSFGGDDHRILV-----SHMKDQHRGIRTGEEF-----	241	
Qy	281	SLINKRYILOLYPHL-----LTEEDVDYNPGLSGCFHDDDDAHAAHNGKPMWIDLNKRY	335	
Db	242	TVTNNSRLLDORQAPAYRETTYQSNNTLAYTGKDLGFVEKLDANAYV-----LEKKRY	293	
Qy	336	ELRAE--WKOPFPQFEALRYELHNRNDVHHDEKAGDAVENFFNNOT--QVARIELRHO---	388	
Db	294	SADDKNDGYAGNVKGNPHNTRITTRGMNFDSRLAE-----QTLCLKYGINYRHQEIK	345	
Qy	389	PIGRLKSGWGOYLGQKS-----SALSATSEAVKQPMLLDNKVQHSYFFGVGEQAN	438	
Db	346	POAFINGKFSIPTTEKKMGQDVAKPADQOAKDKDEALVHSYRLNTPKTDTGAVIEAH	405	
Qy	439	WDN--FTLEGGYVRVEKOKASIRYDKALIDRENYKQPLDPLGAHRQATARSFALSGSNWYFTP	497	
Db	406	EINGFTLTGGLRYDRPKVK--THDGKTVSSNUN-----PSFGVIWOPHE-----HWSFSA	454	
Qy	498	QHKLSLTASHQERLPSTQELVYAHKGKHVANTFVGNKHLNTERSNNIELALGYEGDRWOY	557	
Db	455	SHNY--ASRSPRL--YDALQTHGKR--GIISIGD--GTKAERARNTIEGFNYNEGTFAA	505	
Qy	558	N-----LALYRNEFGNYIYAQTLNDG-----	RGP	581
Db	506	NGRYFWQTIKDALANPQNRHVSAAVREAVNAGYIKNHGYELGASYRTQGLTAKVGVSHSK	565	
Qy	582	KSIDDDSEMKLVRYNQSGADFYGAE-----GEITYFKPTPRYRGVSGDYVRGLKNLP	634	
Db	566	PRFYDTHPKLLSANPE---FGAQVGRWTWASLAYRQNPNLIEIGWRGRYV-----	613	
Qy	635	SLPQREDAYGNRPPIAQDDQNAQPRVPAARLGHKLKASLTDRIDANLDYYRVFPAQNKLYAR	694	
Db	614	-----QKAVGSILVAGQDRNGKLENVVRKGFVN-----DVFANNKPLGK	654	
Qy	695	ETRTPGHHMLNLAGNYRRNTRYGEMWVYVKADNLLNQSVYAHSSFLSDT--PQMGFSFTGG	753	
Db	655	DT-----HNVNLSVN-----NVENTFYFHSQRWNTNLPDVGDRVRLG	692	
Qy	754	VNVKFP	758	
Db	693	VNYKFP	697	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: November 14, 2003, 11:00:43 ; Search time 47 Seconds
(without alignments)
2559.888 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 758
Sequence: 1 MAQTTLKPIVLSILLINTPL.....FLSDTPQMGRSFTGGVNVKF 758

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	758	100.0	758	21	AA18719 A Neisseria mening
2	286	37.7	339	21	AA174562 Neisseria meningit
3	286	37.7	735	21	AA174556 Neisseria meningit
4	286	37.7	758	23	AAU73011 Neisseria meningit
5	192	25.3	758	21	AA174565 Neisseria meningit
6	175	23.1	339	21	AA174563 Neisseria meningit
7	175	23.1	764	21	AA174557 Neisseria meningit
8	175	23.1	764	23	AB578068 Neisseria meningit
9	138	18.2	728	24	ABP80923 N. gonorrhoeae ami

10	138	18.2	758	21	AA174564 Neisseria gonorrh
11	136	17.9	758	21	AA174566 Neisseria meningit
12	117	15.4	758	21	AA174555 Neisseria gonorrh
13	80	10.6	339	21	AA174561 Neisseria gonorrh
14	9	1.2	144	21	AA15254 Arabidopsis thalia
15	9	1.2	145	21	AA15253 Arabidopsis thalia
16	9	1.2	151	21	AA15252 Arabidopsis thalia
17	9	1.2	323	21	AA15252 Arabidopsis thalia
18	9	1.2	323	21	AA15252 Arabidopsis thalia
19	9	1.2	323	21	AA15252 Arabidopsis thalia
20	9	1.2	323	21	AA15252 Arabidopsis thalia
21	9	1.2	323	21	AA15252 Arabidopsis thalia
22	9	1.2	323	21	AA15252 Arabidopsis thalia
23	9	1.2	323	21	AA15252 Arabidopsis thalia
24	9	1.2	1240	24	ABR41659 Human DITP cytosk
25	8	1.1	94	24	ABP58986 Human Mch2 protein
26	8	1.1	99	23	ABR47432 Listeria monocytog
27	8	1.1	115	23	ABP32719 Human ORF192 prot
28	8	1.1	268	22	ABP57898 Drosophila melanog
29	8	1.1	333	22	AAU69473 Human purified sec
30	8	1.1	399	21	AA15255 Formate dehydrogen
31	8	1.1	401	23	AAU99107 Mycobacterium vacc
32	8	1.1	401	23	AAU99108 Mycobacterium vacc
33	8	1.1	401	23	AAU99109 Mycobacterium vacc
34	8	1.1	401	23	AAU99110 Mycobacterium vacc
35	8	1.1	401	23	AAU99111 Mycobacterium vacc
36	8	1.1	401	23	AAU99112 Mycobacterium vacc
37	8	1.1	401	23	AAU99113 Mycobacterium vacc
38	8	1.1	401	23	AAU99114 Mycobacterium vacc
39	8	1.1	401	23	AAU99115 Mycobacterium vacc
40	8	1.1	401	23	AAU99116 Mycobacterium vacc
41	8	1.1	401	23	AAU99117 Mycobacterium vacc
42	8	1.1	401	23	AAU99118 Mycobacterium vacc
43	8	1.1	401	23	AAU99119 Mycobacterium vacc
44	8	1.1	401	23	AAU99120 Mycobacterium vacc
45	8	1.1	401	23	AAU99121 Mycobacterium vacc
46	8	1.1	401	23	AAU99122 Mycobacterium vacc
47	8	1.1	401	23	AAU99123 Mycobacterium vacc
48	8	1.1	430	23	ABP77865 Ancylobacter form
49	8	1.1	446	20	AA15332 Amino acid sequenc
50	7	0.9	10	20	AA15332 Human cytokeatin
51	7	0.9	11	20	AA15332 Cytokeatin 18 apo
52	7	0.9	18	20	AA15332 Cytokeatin 18 apo
53	7	0.9	20	20	AA15332 Amino acid sequenc
54	7	0.9	22	24	ABP71451 Tumour antigen pep
55	7	0.9	22	24	ABP71451 Human breast tumou
56	7	0.9	46	23	ABP78942 Human tumour-relat
57	7	0.9	46	23	ABP78942 Human liver peptid
58	7	0.9	47	22	ABP78942 Peptide #1811 enco
59	7	0.9	47	22	ABP78942 Human brain expres
60	7	0.9	47	22	AA15332 Human bone marrow
61	7	0.9	47	22	AA15332 Peptide #1757 enco
62	7	0.9	47	22	AA15332 Peptide #1835 enco
63	7	0.9	47	22	AA15332 Peptide #1760 enco
64	7	0.9	47	22	AA15332 Human gamma IFN an
65	7	0.9	48	15	AA15332 Sequence ID #639 f
66	7	0.9	51	20	AA15332 Human secreted pro
67	7	0.9	51	22	AA15332 Human immune/haema
68	7	0.9	52	22	AA15332 Propionibacterium
69	7	0.9	56	22	AA15332 Human ORFX protein
70	7	0.9	57	23	ABP7082 Human immune/haema
71	7	0.9	57	22	AA15332 Propionibacterium
72	7	0.9	72	22	AA15332 Human polypeptide
73	7	0.9	73	22	AA15332 Human normal bladd
74	7	0.9	76	24	ABP78539 Propionibacterium
75	7	0.9	82	20	AA15332 Novel human enzyme
76	7	0.9	85	22	AA15332 Human polypeptide
77	7	0.9	105	22	AA15332 Cryptosporidium pa
78	7	0.9	106	22	AA15332 C parvum GP900 pro
79	7	0.9	108	22	AA15332
80	7	0.9	112	21	AA15332
81	7	0.9	112	21	AA15332
82	7	0.9	112	23	ABJ04049

229	7	0.9	520	23	ABP92261	Herbicidally activ	302	7	0.9	4292	22	ABG17060	Novel human diagno
230	7	0.9	524	24	ABP57439	Mycobacterium tube	303	7	0.9	4299	22	ABU52622	Human NOVX protein
231	7	0.9	525	22	AAV72636	Exophiala spinifer	304	7	0.9	4302	17	AAW00870	Polycystic kidney
232	7	0.9	525	23	AAW26386	Exophiala spinifer	305	7	0.9	4302	19	AAW33396	Human PKD1 polypep
233	7	0.9	525	24	ABU07913	Exophiala spinifer	306	7	0.9	4302	19	AAW23830	Human PKD1 protein
234	7	0.9	525	24	ABG75929	Fumonisin degradat	307	7	0.9	4302	21	AAV92719	Human polycystin.
235	7	0.9	533	22	ABG60202	Drosophila melanog	308	7	0.9	4303	17	AAV90302	Polycystic kidney
236	7	0.9	533	22	ABG67019	Drosophila melanog	309	7	0.9	4303	21	AAV95558	Human polycystin-1
237	7	0.9	533	22	ABG67020	Drosophila melanog	310	7	0.9	4303	23	AAE18929	Human PKD1 wild-ty
238	7	0.9	557	22	ABG21200	Novel human diagno	311	7	0.9	4303	23	AAE18932	Human PKD1 mutant
239	7	0.9	562	18	AAW17886	Photorhabdus lumin	312	7	0.9	4303	23	AAE18934	Human PKD1 mutant
240	7	0.9	562	19	AAW56545	Toxin TcaBii, enco	313	7	0.9	4303	23	AAE18936	Human PKD1 mutant
241	7	0.9	593	23	ABBS3293	Human polypeptide	314	7	0.9	4303	23	AAE18937	Human PKD1 mutant
242	7	0.9	595	23	AAE16337	Human POLY1 protei	315	7	0.9	4303	23	AAE18938	Human PKD1 mutant
243	7	0.9	595	23	AAE16338	Human POLY2 protei	316	7	0.9	4303	23	AAE18939	Human PKD1 mutant
244	7	0.9	595	23	AAE16339	Human POLY3 protei	317	7	0.9	4303	23	AAE18940	Human PKD1 mutant
245	7	0.9	596	22	AAE67544	Amino acid sequenc	318	7	0.9	4303	23	AAE18941	Human PKD1 mutant
246	7	0.9	596	23	AAE22723	Human cardiac myos	319	7	0.9	4303	23	AAE18942	Human PKD1 mutant
247	7	0.9	596	23	AAE22847	Human cardiac myos	320	7	0.9	4303	23	AAE18943	Human PKD1 mutant
248	7	0.9	596	23	AAE22848	Human cardiac myos	321	7	0.9	4303	23	AAE18944	Human PKD1 mutant
249	7	0.9	596	23	AAE22849	Human cardiac myos	322	7	0.9	4304	23	AAE18951	Human PKD1 mutant
250	7	0.9	596	23	AAE22850	Human cardiac myos	323	7	0.9	4304	23	AAE18951	Human PKD1 mutant
251	7	0.9	596	23	AAE16260	Human kinase PKIN-	324	7	0.9	4339	17	AAE18951	Polycystic kidney
252	7	0.9	605	23	AAE16260	Human POLY4 protei	325	7	0.9	4339	17	AAE18951	Novel human diagno
253	7	0.9	612	22	AAU03521	Herbicidally activ	326	7	0.9	4977	22	ABG17057	Novel human diagno
254	7	0.9	632	22	ABG07302	Human protein kina	327	7	0.9	6685	22	ABG23030	Novel human diagno
255	7	0.9	648	22	AAU02193	Novel human diagno	328	6	0.8	6	18	AAW29947	Cysteine protease
256	7	0.9	648	22	ABP35626	Transcriptional ac	329	6	0.8	7	18	AAW32279	Cysteine protease
257	7	0.9	680	23	ABP64766	Fungal ZBC protein	330	6	0.8	7	18	AAW29946	Cysteine protease
258	7	0.9	680	23	ABP64766	Human protein SEQ	331	6	0.8	7	18	AAW29948	Cysteine protease
259	7	0.9	685	22	ABBS8231	Drosophila melanog	332	6	0.8	7	18	AAW29949	Cysteine protease
260	7	0.9	720	23	AAU84293	Human endometrial	333	6	0.8	7	21	AAE21115	Synthetic caspase
261	7	0.9	786	22	ABBS9818	Drosophila melanog	334	6	0.8	7	22	AAE67724	Amino acid sequenc
262	7	0.9	802	23	ABP49266	Listeria monocytog	335	6	0.8	7	22	AAE67724	Caspase-2 protease
263	7	0.9	814	23	AAE17586	M catarrhalis MCAI	336	6	0.8	8	17	AAW96818	N.gonorrhoeae MS11
264	7	0.9	818	22	AAE19945	Moraxella catarrha	337	6	0.8	8	17	AAW29950	Cysteine protease
265	7	0.9	852	22	AAW51656	Corynebacterium gl	338	6	0.8	8	18	AAW29951	Cysteine protease
266	7	0.9	852	22	AAW51656	C glutamicum prote	339	6	0.8	8	24	ABP74863	Proteome analysis
267	7	0.9	930	23	AAU79718	Trypanosoma brucei	340	6	0.8	8	24	ABP75079	Proteome analysis
268	7	0.9	970	23	ABBS69643	Drosophila melanog	341	6	0.8	9	20	AAV55507	HLA binding plu-1
269	7	0.9	970	23	ABP27418	Streptococcus poly	342	6	0.8	9	20	AAV55507	HLA binding plu-1
270	7	0.9	1018	22	AAU33690	Pseudomonas aerugi	343	6	0.8	9	20	AAV55610	HLA binding plu-1
271	7	0.9	1035	22	ABBS8404	Drosophila melanog	344	6	0.8	9	22	AAV55610	HLA class I bindin
272	7	0.9	1039	22	ABBS63666	Drosophila melanog	345	6	0.8	9	22	AAV55610	Influenza virus im
273	7	0.9	1043	22	ABBS60430	Drosophila melanog	346	6	0.8	10	23	AAW49904	Human D40 associat
274	7	0.9	1189	18	AAW17884	Photorhabdus lumin	347	6	0.8	11	13	AAW25097	bGRF produg analo
275	7	0.9	1189	19	AAW56543	Toxin TcaB, encode	348	6	0.8	11	21	AAO22844	Angiogenic treatme
276	7	0.9	1311	20	AAW52197	Precis coenia (but	349	6	0.8	11	21	AAO22844	Angiogenic treatme
277	7	0.9	1311	20	AAW72971	Butterfly patched	350	6	0.8	11	23	ABP54036	Radopharmaceutica
278	7	0.9	1311	22	ABBS67156	Butterfly patched	351	6	0.8	11	23	AAU79102	Synthetic peptide
279	7	0.9	1311	21	AAU79573	Novel human diagno	352	6	0.8	11	24	ABG73009	Radolabelled Vitr
280	7	0.9	1511	21	AAU38528	Novel human diagno	353	6	0.8	14	21	AAV99222	HLA class II bindi
281	7	0.9	1512	22	ABG20454	Novel human secret	354	6	0.8	14	22	AAW53154	Human conservative
282	7	0.9	1525	22	AAU30437	Arabidopsis thalia	355	6	0.8	14	22	AAW53156	Human conservative
283	7	0.9	1544	21	AAU38527	Arabidopsis thalia	356	6	0.8	15	17	AAW07276	Amphiphilic antimi
284	7	0.9	1547	21	AAU38526	Novel human diagno	357	6	0.8	15	21	AAE27805	Human secreted pro
285	7	0.9	1570	22	ABG23840	Human PKD1 truncat	358	6	0.8	16	23	ABP79346	Human ovary specif
286	7	0.9	1653	23	AAE18945	Human PKD1 truncat	359	6	0.8	16	24	ABP79346	Amino acid sequenc
287	7	0.9	1721	19	AAW48299	Cryptosporidium pa	360	6	0.8	17	24	AAE32989	Human sperm activa
288	7	0.9	1721	21	AAE11727	Portion of Cryptos	361	6	0.8	19	21	AAV79455	Nematode kynurenin
289	7	0.9	1721	21	ABU04045	C parvum gp900 pro	362	6	0.8	19	22	AAE23897	PIR-A1 receptor pe
290	7	0.9	1837	21	AAE11726	Cryptosporidium pa	363	6	0.8	21	22	AAU08479	Peptide #1 inhibit
291	7	0.9	1837	23	ABJ04044	C parvum gp900 pro	364	6	0.8	21	22	AAU08480	Peptide #2 inhibit
292	7	0.9	1960	23	AAE18946	Human PKD1 truncat	365	6	0.8	21	22	AAU08486	VEGFR-3 binding ep
293	7	0.9	2039	23	AAE18947	Human PKD1 truncat	366	6	0.8	22	22	ABG07762	Novel human diagno
294	7	0.9	2089	17	AAW08333	Cyclotella cryptic	367	6	0.8	22	23	ABG76660	Fruit fly SOUP1-sp
295	7	0.9	2402	23	AAE18948	Human PKD1 truncat	368	6	0.8	24	21	AAV56833	Apple LRPXmi LRR s
296	7	0.9	2424	22	ABBS8924	Drosophila melanog	369	6	0.8	24	23	ABG76661	Fruit fly SOUP1-sp
297	7	0.9	2430	23	AAE18950	Human PKD1 truncat	370	6	0.8	26	21	AAV68638	Amino acid sequenc
298	7	0.9	2558	23	AAE18949	Human PKD1 truncat	371	6	0.8	28	19	AAW60193	Bacteriophage spo2
299	7	0.9	3001	23	AAE18944	Human PKD1 truncat	372	6	0.8	29	22	ABG55564	Human liver peptid
300	7	0.9	3433	18	AAW22017	Utrrophin. Homo sa	373	6	0.8	29	22	ABB40308	Peptide #7814 enco
301	7	0.9	3858	23	ABP69424	Human polypeptide	374	6	0.8	29	22	ABB24711	Protein #6710 enco

375	6	0.8	29	22	AAM61106	Human brain expres	448	6	0.8	55	22	ABB17611	Human nervous syst
376	6	0.8	29	22	AAM73814	Human bone marrow	449	6	0.8	55	22	AAO23957	Human EST encoded
377	6	0.8	29	22	AAM20110	Peptide #6544 enco	450	6	0.8	55	23	AAO21812	Lung-specific anin
378	6	0.8	29	22	AAM34000	Peptide #8037 enco	451	6	0.8	56	18	AAO27943	Staphylococcus aur
379	6	0.8	29	23	ABG43703	Human peptide enco	452	6	0.8	56	20	AAO48293	Human prostate can
380	6	0.8	30	20	AAM88907	Polypeptide fragme	453	6	0.8	56	21	ABG58853	Arabidopsis thalia
381	6	0.8	30	22	ABG50976	Human secreted pro	454	6	0.8	57	21	ABG39354	Arabidopsis thalia
382	6	0.8	31	22	AAO11273	Human polypeptide	455	6	0.8	57	22	AAO67416	Propionibacterium
383	6	0.8	31	23	ABG76338	Haemophilus influe	456	6	0.8	57	22	ABG52539	Escherichia coli p
384	6	0.8	32	21	AAO44209	Fusion protein com	457	6	0.8	57	23	ABP32272	Human ORF1245 prot
385	6	0.8	32	24	ABR01256	Human gene 310-enc	458	6	0.8	57	24	ABU02022	S. pneumoniae type
386	6	0.8	33	24	ABP99766	Human secreted pro	459	6	0.8	58	21	ABG51676	Human secreted pro
387	6	0.8	34	20	AAO11749	Human 5' EST secre	460	6	0.8	58	21	AAO1340	Propionibacterium
388	6	0.8	34	21	ABG45260	Human secreted pro	461	6	0.8	58	22	AAO61398	Human 5' EST secre
389	6	0.8	36	22	ABG27767	Novel human diagno	462	6	0.8	59	20	AAO12900	Hepatitis GB virus
390	6	0.8	37	22	ABG52653	Human liver peptid	463	6	0.8	59	21	ABO9313	Propionibacterium
391	6	0.8	37	22	AAO08930	Human polypeptide	464	6	0.8	59	22	AAO67771	Arabidopsis thalia
392	6	0.8	37	22	ABG31225	Peptide #5262 enco	465	6	0.8	60	21	ABG18583	Novel human diagno
393	6	0.8	37	23	ABG40731	Human peptide enco	466	6	0.8	60	22	ABG24452	Human immune/haema
394	6	0.8	38	22	AAO85197	Human immune/haema	467	6	0.8	60	22	AAO86975	Human ORFX protein
395	6	0.8	38	22	AAO05509	Human polypeptide	468	6	0.8	60	23	ABP00659	Arabidopsis thalia
396	6	0.8	40	22	ABG01044	Novel human diagno	469	6	0.8	61	21	AAO15330	Propionibacterium
397	6	0.8	41	22	ABG55797	Human liver peptid	470	6	0.8	61	22	AAU44047	Human immune/haema
398	6	0.8	41	22	AAO74043	Human bone marrow	471	6	0.8	61	22	AAO86042	Human complement t
399	6	0.8	41	23	ABG43939	Human peptide enco	472	6	0.8	62	13	AAO28540	CR1 SCR-2. Homo s
400	6	0.8	42	22	ABO3641	Human musculoskele	473	6	0.8	62	16	AAO72333	Human breast cance
401	6	0.8	42	24	ABU12935	Novel human muscul	474	6	0.8	62	18	AAO58758	Human ORFX protein
402	6	0.8	43	22	ABG55121	Human liver peptid	475	6	0.8	62	18	AAO27797	Arabidopsis thalia
403	6	0.8	43	22	ABG39981	Peptide #7487 enco	476	6	0.8	62	22	AAU54008	Propionibacterium
404	6	0.8	43	22	AAO60728	Human brain expres	477	6	0.8	62	22	AAU61784	Propionibacterium
405	6	0.8	43	22	AAO73399	Human bone marrow	478	6	0.8	62	22	AAO50409	Human immune/haema
406	6	0.8	43	22	ABG64622	Human secreted pro	479	6	0.8	62	22	AAO1681	Gene 24 human secr
407	6	0.8	43	23	ABG43258	Human peptide enco	480	6	0.8	62	22	ABG29527	Human CR1 compleme
408	6	0.8	44	24	ABP77445	N. gonorrhoeae ami	481	6	0.8	62	23	ABP51839	Human colon specif
409	6	0.8	45	21	AAO44208	Protein comprising	482	6	0.8	63	20	AAO59817	Human normal ovari
410	6	0.8	45	23	ABG69811	Human secretory pr	483	6	0.8	63	21	ABG27877	Sequence homologou
411	6	0.8	45	23	ABP61398	Breast specific po	484	6	0.8	64	22	ABO50978	Novel bone marrow
412	6	0.8	46	20	AAO25825	Human secreted pro	485	6	0.8	64	22	AAO59592	Human testicular a
413	6	0.8	46	23	ABP26164	Streptococcus poly	486	6	0.8	64	22	ABG59552	Propionibacterium
414	6	0.8	46	23	ABP68165	Streptococcus poly	487	6	0.8	64	22	AAO59525	Human reproductive
415	6	0.8	47	21	ABG58669	Arabidopsis thalia	488	6	0.8	64	23	ABG53859	Lactococcus lactis
416	6	0.8	47	22	ABG56104	Human liver peptid	489	6	0.8	64	24	AAO34739	Streptomyces tuber
417	6	0.8	47	22	ABG01303	Novel human diagno	490	6	0.8	64	24	AAO34770	Streptomyces tuber
418	6	0.8	47	22	ABO40656	Peptide #8162 enco	491	6	0.8	65	21	AAO37751	Arabidopsis thalia
419	6	0.8	47	22	ABO24909	Protein #6908 enco	492	6	0.8	65	21	AAO39353	Human liver peptid
420	6	0.8	47	22	AAO61516	Human brain expres	493	6	0.8	65	22	ABG58958	Arabidopsis thalia
421	6	0.8	47	22	AAO74305	Human bone marrow	494	6	0.8	65	22	ABG43580	Human ORFX protein
422	6	0.8	47	23	ABG34417	Peptide #8454 enco	495	6	0.8	65	22	ABG26539	Peptide #11086 enco
423	6	0.8	47	23	ABG44232	Human peptide enco	496	6	0.8	65	22	ABG4517	Protein #8538 enco
424	6	0.8	48	21	AAO30478	Amino acid sequenc	497	6	0.8	65	22	AAO77325	Human brain expres
425	6	0.8	48	21	AAO30479	Amino acid sequenc	498	6	0.8	65	22	AAO21251	Human bone marrow
426	6	0.8	49	22	ABG54660	Human liver peptid	499	6	0.8	65	22	AAO21251	Peptide #7685 enco
427	6	0.8	49	22	ABG57145	Human liver peptid	500	6	0.8	65	22	AAO37477	Peptide #11514 enc
428	6	0.8	49	22	ABG39613	Peptide #7213 enco	501	6	0.8	65	23	ABG46340	Human peptide enco
429	6	0.8	49	22	ABO41705	Peptide #7211 enco	502	6	0.8	65	23	ABO41201	Human ovarian anti
430	6	0.8	49	22	AAO60317	Peptide #9211 enco	503	6	0.8	65	23	ABO6519	Human ORFX protein
431	6	0.8	49	22	AAO62579	Human brain expres	504	6	0.8	66	22	ABG55620	Human liver peptid
432	6	0.8	49	22	AAO72951	Human brain expres	505	6	0.8	66	22	AAO41059	Propionibacterium
433	6	0.8	49	22	AAO75389	Human bone marrow	506	6	0.8	66	22	AAO73869	Human bone marrow
434	6	0.8	49	22	AAO33176	Human bone marrow	507	6	0.8	66	22	AAO85976	Human immune/haema
435	6	0.8	49	22	AAO35502	Peptide #7213 enco	508	6	0.8	66	22	AAO34051	Peptide #8088 enco
436	6	0.8	49	23	ABG42788	Peptide #9539 enco	509	6	0.8	66	23	ABG43757	Human peptide enco
437	6	0.8	50	22	AAO56264	Human peptide enco	510	6	0.8	67	22	ABG65190	Propionibacterium
438	6	0.8	50	22	AAO65180	Propionibacterium	511	6	0.8	67	22	ABG16697	Novel human diagno
439	6	0.8	50	22	AAO20214	Propionibacterium	512	6	0.8	67	22	ABO10780	Human ovarian and/
440	6	0.8	52	22	AAO05589	Human novel endocr	513	6	0.8	67	22	AAO94443	Human reproductive
441	6	0.8	53	18	AAO11234	Human ORFX protein	514	6	0.8	67	23	AAO20978	S. lugdunensis von
442	6	0.8	53	21	AAO15930	S. pneumoniae prot	515	6	0.8	67	23	AAO20978	S. lugdunensis von
443	6	0.8	53	23	ABP34506	Arabidopsis thalia	516	6	0.8	68	23	ABP40328	Staphylococcus epi
444	6	0.8	54	22	AAO21263	Human ORFX3479 prot	517	6	0.8	68	24	AAO30259	Human LP289 protei
445	6	0.8	54	22	AAO08422	Interleukin-2 homo	518	6	0.8	69	20	AAO48374	Human prostate can
446	6	0.8	54	23	ABP08906	Polypeptide encode	519	6	0.8	69	22	AAO58582	Propionibacterium
447	6	0.8	55	21	AAO56900	Human ORFX protein	520	6	0.8	69	22	AAO51061	Propionibacterium
						Human prostate can						AAO61858	Propionibacterium

521	6	0.8	69	22	ABG29282	Novel human diagno	594	6	0.8	88	22	AAO13725	Human polypeptide
522	6	0.8	70	22	AAU62960	Propionibacterium	595	6	0.8	88	23	ABF04898	Human ORFX protein
523	6	0.8	70	23	ABP35014	Human ORF987 prot	596	6	0.8	89	21	ABG40445	Arabidopsis thalia
524	6	0.8	71	21	ABP35851	Human colon cancer	597	6	0.8	90	22	ABG18696	Novel human diagno
525	6	0.8	71	22	AAU54956	Propionibacterium	598	6	0.8	90	22	ABG27030	Novel human diagno
526	6	0.8	71	23	AAU63746	Propionibacterium	599	6	0.8	90	22	AAU22539	Novel human colon
527	6	0.8	71	23	ABP06343	Human ORFX protein	600	6	0.8	90	22	AAU22539	Human digestive sy
528	6	0.8	72	20	AAU59922	Human myometrium t	601	6	0.8	91	21	AAU57594	Arabidopsis thalia
529	6	0.8	72	21	AAU07879	Arabidopsis thalia	602	6	0.8	91	21	AAU57594	Arabidopsis thalia
530	6	0.8	72	21	AAU32901	Arabidopsis thalia	603	6	0.8	91	21	AAU57594	Arabidopsis thalia
531	6	0.8	72	21	AAU45329	Arabidopsis thalia	604	6	0.8	91	21	AAU57594	Arabidopsis thalia
532	6	0.8	72	22	AAU44084	Propionibacterium	605	6	0.8	91	22	AAU29630	Novel human diagno
533	6	0.8	72	22	ABG11779	Novel human diagno	606	6	0.8	91	22	AAU29630	Novel human diagno
534	6	0.8	72	23	ABP07293	Human ORFX protein	607	6	0.8	91	22	AAU09573	Human polypeptide
535	6	0.8	72	23	ABP53109	Human ORF15 protei	608	6	0.8	91	23	ABG64776	Human albumin fusi
536	6	0.8	73	15	AAU49660	Sequence of Heliot	609	6	0.8	91	23	ABG64776	Human albumin fusi
537	6	0.8	73	20	AAU29076	T. gondii immunoge	610	6	0.8	91	23	AAO17185	Human secreted pro
538	6	0.8	73	22	AAU42135	Propionibacterium	611	6	0.8	91	23	AAO17185	Human secreted pro
539	6	0.8	73	22	ABG04681	Novel human diagno	612	6	0.8	92	20	AAU11433	Human 5' EST secre
540	6	0.8	73	22	ABP03925	Human musculoskele	613	6	0.8	92	20	AAU11433	Human 5' EST secre
541	6	0.8	73	22	AAU25547	Novel human diagno	614	6	0.8	92	21	AAU40444	Arabidopsis thalia
542	6	0.8	73	24	ABU13219	T. gondii immunoge	615	6	0.8	92	22	ABG15046	Novel human diagno
543	6	0.8	74	22	AAU55320	Novel human muscul	616	6	0.8	92	22	ABG16016	Novel human diagno
544	6	0.8	74	22	ABU10404	Propionibacterium	617	6	0.8	92	23	ABP53320	Human cell signall
545	6	0.8	74	22	AAU18095	Human cDNA SQ ID	618	6	0.8	92	23	ABP53320	Human cell signall
546	6	0.8	74	22	AAU16992	Novel human uterin	619	6	0.8	93	23	ABP42406	Human ORFX protein
547	6	0.8	74	22	AAU19911	Human novel secret	620	6	0.8	94	21	AAU17793	Human ovarian anti
548	6	0.8	74	23	ABP66991	Novel human calciu	621	6	0.8	94	21	AAU17793	Human ovarian anti
549	6	0.8	74	23	ABU05722	Human polypeptide	622	6	0.8	94	21	AAU17793	Human ovarian anti
550	6	0.8	74	23	ABU40775	Novel human protei	623	6	0.8	94	21	AAU17793	Human ovarian anti
551	6	0.8	75	22	AAU47596	Staphylococcus epi	624	6	0.8	94	22	AAU42596	Arabidopsis thalia
552	6	0.8	75	22	AAU51458	Propionibacterium	625	6	0.8	94	22	ABU16469	Arabidopsis thalia
553	6	0.8	75	22	AAU52259	Propionibacterium	626	6	0.8	94	22	AAU02044	Human nervous syst
554	6	0.8	75	22	ABG11849	Propionibacterium	627	6	0.8	94	22	AAU02044	Human nervous syst
555	6	0.8	75	22	ABG17582	Novel human diagno	628	6	0.8	94	22	AAU08326	Human polypeptide
556	6	0.8	76	21	AAU26731	Human colon cancer	629	6	0.8	95	22	AAU58826	Human polypeptide
557	6	0.8	76	22	ABG18691	Zea mays protein f	630	6	0.8	95	22	AAU58826	Human ORFX protein
558	6	0.8	76	22	AAU83699	Novel human diagno	631	6	0.8	95	22	AAU50760	Propionibacterium
559	6	0.8	76	22	AAU91147	Human immune/haema	632	6	0.8	95	22	AAU50760	Propionibacterium
560	6	0.8	76	22	ABP51882	Human immune/haema	633	6	0.8	95	22	AAU50760	Propionibacterium
561	6	0.8	77	22	ABG58654	Human colon specif	634	6	0.8	96	24	ABP77579	Human immune/haema
562	6	0.8	77	22	AAU56370	Human liver peptid	635	6	0.8	96	24	ABP77579	Human immune/haema
563	6	0.8	77	22	AAU59597	Propionibacterium	636	6	0.8	97	21	AAU40354	Human ORFX ORF118
564	6	0.8	77	22	AAU67969	Propionibacterium	637	6	0.8	97	21	AAU40354	Human ORFX ORF118
565	6	0.8	77	22	AAU76993	Propionibacterium	638	6	0.8	97	22	ABG22936	Arabidopsis thalia
566	6	0.8	77	22	AAU84410	Human bone marrow	639	6	0.8	97	22	ABG22936	Arabidopsis thalia
567	6	0.8	77	22	AAU37101	Human immune/haema	640	6	0.8	97	22	ABG22936	Arabidopsis thalia
568	6	0.8	77	24	ABU18362	Peptide #1138 enc	641	6	0.8	97	22	ABG22936	Arabidopsis thalia
569	6	0.8	78	19	AAU63047	Breast specific re	642	6	0.8	97	23	ABG60032	Equine influenza v
570	6	0.8	79	22	AAU74849	Streptococcus uber	643	6	0.8	97	23	ABG60032	Equine influenza v
571	6	0.8	80	22	AAU75529	Human colon cancer	644	6	0.8	97	23	ABP04327	Human DITP polype
572	6	0.8	80	22	ABP03379	Human colon cancer	645	6	0.8	97	23	ABP04327	Staphylococcus epi
573	6	0.8	81	16	AAU70158	Human ORFX protein	646	6	0.8	98	22	ABP95721	Human ORFX protein
574	6	0.8	81	22	AAU92596	Streptococcus pneu	647	6	0.8	98	22	ABP95721	Human testicular a
575	6	0.8	81	22	AAU92596	C glutamicum prote	648	6	0.8	98	22	AAU62088	Propionibacterium
576	6	0.8	81	23	AAU72564	Corynebacterium gl	649	6	0.8	98	22	AAU62088	Propionibacterium
577	6	0.8	81	23	AAU72564	Human ORFX protein	650	6	0.8	98	22	AAU62088	Propionibacterium
578	6	0.8	82	24	AAU01133	S. pneumoniae type	651	6	0.8	98	22	AAU62088	Propionibacterium
579	6	0.8	83	22	ABG25261	Novel human diagno	652	6	0.8	98	22	AAU62088	Propionibacterium
580	6	0.8	83	23	ABP09336	Human ORFX protein	653	6	0.8	98	22	AAU62088	Propionibacterium
581	6	0.8	83	23	AAU43363	Propionibacterium	654	6	0.8	98	22	AAU62088	Propionibacterium
582	6	0.8	85	22	AAU80328	Human haematologic	655	6	0.8	98	22	AAU62088	Propionibacterium
583	6	0.8	85	23	ABP06148	Human ORFX protein	656	6	0.8	98	22	AAU62088	Propionibacterium
584	6	0.8	86	22	AAU52823	Propionibacterium	657	6	0.8	98	22	AAU62088	Propionibacterium
585	6	0.8	86	22	ABP15796	Human nervous syst	658	6	0.8	98	22	AAU62088	Propionibacterium
586	6	0.8	86	23	ABP35536	Human ORF4509 prot	659	6	0.8	98	22	AAU62088	Propionibacterium
587	6	0.8	87	22	AAU44455	Propionibacterium	660	6	0.8	98	22	AAU62088	Propionibacterium
588	6	0.8	87	22	AAU53060	Propionibacterium	661	6	0.8	98	22	AAU62088	Propionibacterium
589	6	0.8	88	21	AAU43048	Human ORFX ORF2812	662	6	0.8	98	22	AAU62088	Propionibacterium
590	6	0.8	88	21	AAU13004	Arabidopsis thalia	663	6	0.8	98	22	AAU62088	Propionibacterium
591	6	0.8	88	21	AAU21579	Arabidopsis thalia	664	6	0.8	98	22	AAU62088	Propionibacterium
592	6	0.8	88	21	AAU37749	Arabidopsis thalia	665	6	0.8	98	22	AAU62088	Propionibacterium
593	6	0.8	88	21	AAU52968	Arabidopsis thalia	666	6	0.8	98	22	AAU62088	Propionibacterium

813	6	0.8	123	21	AAG44953	Zea mays protein f	886	139	22	ABG03688	Novel human diagno
814	6	0.8	123	22	AAO06597	Human polypeptide	887	139	22	AAE03783	Human gene 1 encod
815	6	0.8	123	22	ABU05142	Pig galanin relate	888	139	22	AAE01703	Human gene 4 encod
816	6	0.8	123	23	ABU10562	Galanin-like pepi	889	139	23	ABG63979	Human albumin fusi
817	6	0.8	124	21	AAG13062	Arabidopsis thalia	890	139	23	ABG65498	Human albumin fusi
818	6	0.8	124	21	AAG16950	Arabidopsis thalia	891	140	22	ABG71829	Infectious salmon
819	6	0.8	124	21	AAG21859	Arabidopsis thalia	892	140	24	ABB84620	Human wild-type VE
820	6	0.8	124	23	ABP00672	Human ORFX protein	893	140	24	ABG73749	Human VEGF-C monom
821	6	0.8	125	13	ABP24512	Prod. of UCD-2 of	894	141	20	AAU37022	Protein which is s
822	6	0.8	125	14	ABR24512	Feline infectious	895	141	21	AAU44952	Zea mays protein f
823	6	0.8	125	14	ABR24519	Arabidopsis thalia	896	142	20	AAW85738	Polypeptide with t
824	6	0.8	125	22	AAU19592	Propionibacterium	897	142	20	AAW83900	Antigen 2 from clu
825	6	0.8	126	21	AAU32296	Arabidopsis thalia	898	142	21	AAU57323	Arabidopsis thalia
826	6	0.8	126	21	AAU57751	Arabidopsis thalia	899	142	22	AAU25727	Breast cancer-asso
827	6	0.8	126	21	AAU60251	Arabidopsis thalia	900	142	22	AAU05584	Human breast cance
828	6	0.8	126	22	AAU08408	Polypeptide encode	901	142	23	ABB89954	Human polypeptide
829	6	0.8	126	22	AAU08409	Polypeptide encode	902	142	23	AAU35954	Extended human sec
830	6	0.8	126	22	AAU08410	Polypeptide encode	903	144	22	ABB95880	Human testicular a
831	6	0.8	126	22	AAU08412	Polypeptide encode	904	144	22	AAU43167	Propionibacterium
832	6	0.8	126	22	AAU08413	Polypeptide encode	905	144	22	AAU95180	Human reproductive
833	6	0.8	126	22	AAU08414	Polypeptide encode	906	144	22	AAU43092	Human ORFX ORF2856
834	6	0.8	127	21	AAU41278	Human ORFX ORF1042	907	145	22	AAO06479	Human polypeptide
835	6	0.8	127	22	AAU08405	Polypeptide encode	908	145	23	ABB80787	Magnetic particle
836	6	0.8	127	23	ABP03223	Human ORFX protein	909	146	21	AAU16299	Eucalyptus grandis
837	6	0.8	127	23	ABU10027	Insulin-like growt	910	146	21	AAU16299	Arabidopsis thalia
838	6	0.8	128	21	AAU58628	Arabidopsis thalia	911	146	21	AAU16299	Arabidopsis thalia
839	6	0.8	128	22	AAU08415	Polypeptide encode	912	146	22	AAU40268	Arabidopsis thalia
840	6	0.8	128	22	AAU08416	Polypeptide encode	913	146	22	AAU93192	Human protein sequ
841	6	0.8	128	22	AAU08417	Polypeptide encode	914	146	22	AAU94383	Human protein sequ
842	6	0.8	128	22	AAU08418	Polypeptide encode	915	146	22	AAU94403	Human protein sequ
843	6	0.8	128	22	AAU08419	Polypeptide encode	916	146	22	AAU94983	Human protein sequ
844	6	0.8	128	22	AAU08421	Polypeptide encode	917	146	22	AAU48932	MuCA protein. Uni
845	6	0.8	129	21	AAU46659	Arabidopsis thalia	918	147	22	AAU48932	Drosophila melanog
846	6	0.8	129	21	AAU00223	Human secreted pro	919	147	23	ABG701594	Human prey protein
847	6	0.8	129	22	ABG07629	Novel human diagno	920	147	23	ABB89291	Human polypeptide
848	6	0.8	129	22	AAU08406	Polypeptide encode	921	148	21	AAU09454	Arabidopsis thalia
849	6	0.8	130	20	AAU01485	ApM1 protein fragm	922	148	21	AAU52382	Arabidopsis thalia
850	6	0.8	130	20	AAU01486	AdipoQ protein fra	923	148	22	AAU021241	Novel interleukin-
851	6	0.8	130	20	AAU01487	acrp30 protein fra	924	149	21	AAU40220	Human secreted pro
852	6	0.8	130	21	AAU04561	Arabidopsis thalia	925	149	24	ABF75640	Human secretory po
853	6	0.8	131	21	AAU57324	Arabidopsis thalia	926	150	19	AAW83187	Staphylococcus aur
854	6	0.8	131	22	AAU23233	Novel human enzyme	927	151	21	AAU16268	Pinus radiata AGP
855	6	0.8	131	22	ABU51489	Helicobacter pylor	928	151	21	AAU18952	Zea mays protein f
856	6	0.8	131	23	ABP08216	Human ORFX protein	929	151	21	AAU44246	Arabidopsis thalia
857	6	0.8	132	17	ABR92117	Human Pac-1. Homo	930	151	21	AAU59845	Arabidopsis thalia
858	6	0.8	132	21	AAU07123	Arabidopsis thalia	931	151	21	AAU60285	Arabidopsis thalia
859	6	0.8	132	22	AAU078432	Human protein SEQ	932	151	22	ABU16240	Novel human diagno
860	6	0.8	132	22	AAU04159	Human polypeptide	933	151	22	ABU12182	Human PRO4990 poly
861	6	0.8	133	21	AAU86482	Human gene 54-enco	934	151	24	ABU66580	Human secreted/tra
862	6	0.8	133	22	AAU32290	Novel human secret	935	151	24	ABU66856	Human secreted and
863	6	0.8	133	22	AAU010761	Human polypeptide	936	151	24	ABU59661	Novel secreted aur
864	6	0.8	134	20	AAU36108	Extended human sec	937	152	19	AAU77504	Human endometrium
865	6	0.8	134	21	AAU13061	Arabidopsis thalia	938	152	20	AAU60208	Extended human sec
866	6	0.8	134	21	AAU15718	Arabidopsis thalia	939	152	20	AAU59363	Human SEEK1 protei
867	6	0.8	134	21	AAU21858	Arabidopsis thalia	940	152	23	ABP53108	Human SEEK1 wild t
868	6	0.8	134	21	AAU44247	Arabidopsis thalia	941	152	23	ABP53109	Human SEEK-1 prote
869	6	0.8	134	21	AAU54793	Arabidopsis thalia	942	152	23	AAU79581	Human SEEK-1 prote
870	6	0.8	135	22	AAU62140	Propionibacterium	943	152	23	AAU79760	Bcl-XL. Homo sapi
871	6	0.8	135	22	ABG08550	Novel human diagno	944	152	24	ABP38913	Staphylococcus epi
872	6	0.8	135	22	AAU08879	Human polypeptide	945	153	23	ABP38913	M. tuberculosis lu
873	6	0.8	135	22	AAU20661	Human secreted pro	946	154	22	AAU50374	Propionibacterium
874	6	0.8	136	21	AAU07644	Arabidopsis thalia	947	154	22	AAU50374	Human protein sequ
875	6	0.8	136	21	AAU49504	Arabidopsis thalia	948	154	22	AAU50374	Amino acid sequenc
876	6	0.8	136	21	AAU77977	A. thaliana enviro	949	155	18	AAU27977	LOX-1 isoform 9.
877	6	0.8	136	22	ABG16919	Novel human diagno	950	155	24	AAU79880	Arabidopsis thalia
878	6	0.8	137	17	ABR86948	PRP35 traA fusion	951	156	21	AAU11340	Novel human diagno
879	6	0.8	137	22	ABG02021	Novel human diagno	952	156	22	AAU24219	Human immune/naema
880	6	0.8	137	22	ABG24915	Novel human diagno	953	156	22	AAU89077	Protein encoded by
881	6	0.8	137	22	AAU81505	S. epidermidis ope	954	157	20	AAU89801	C glutamic prote
882	6	0.8	137	24	ABP76268	Human GENSEP prote	955	157	22	AAU89846	Human BAFF recepto
883	6	0.8	137	24	ABU99359	Amino acid sequenc	956	157	22	AAU60700	Corynebacterium gl
884	6	0.8	138	17	AAU92524	Pyrodictium occult	957	158	17	AAU94674	Flea saliva protei
885	6	0.8	138	21	AAU36064	Zea mays protein f	958	158	23	AAU05420	M. tuberculosis an

959	6	0.8	159	17	AAW03545	Partial ORF-6 prot
960	6	0.8	159	23	ABU51129	Helicobacter pylori
961	6	0.8	159	23	ABB80777	A. salmonicida typ
962	6	0.8	160	21	AA010791	M. avium lumazine
963	6	0.8	160	22	AA005551	Human polypeptide
964	6	0.8	160	23	AAU96863	Human eIF-2 associ
965	6	0.8	161	21	AA058843	Breast and ovarian
966	6	0.8	161	23	ABP53605	Myo-inositol oxyge
967	6	0.8	161	24	ABU11785	Human MDDT polypep
968	6	0.8	162	21	AA0344252	Human secreted pro
969	6	0.8	162	22	AA098488	Papio hamadryas ol
970	6	0.8	163	21	AA020224	Arabidopsis thalia
971	6	0.8	163	21	AA045249	Arabidopsis thalia
972	6	0.8	163	22	AAU44160	Propionibacterium
973	6	0.8	163	22	AAU59359	Propionibacterium
974	6	0.8	163	23	ABP69434	Human polypeptide
975	6	0.8	163	23	ABB08160	Human cytoskeleton
976	6	0.8	163	23	ABP26273	Streptococcus poly
977	6	0.8	164	21	AA010812	Synechocystis sp l
978	6	0.8	164	22	ABG01387	Novel human diagno
979	6	0.8	164	22	AAU68931	Human protease dom
980	6	0.8	165	15	AA047459	Truncated xylanase
981	6	0.8	165	15	AA051460	19kD fibrinogen bi
982	6	0.8	165	15	AA051462	19kD fibrinogen bi
983	6	0.8	165	20	AA049029	Cloning vector PRI
984	6	0.8	166	20	AA053034	Cloning vector PRI
985	6	0.8	166	22	AAU54895	Propionibacterium
986	6	0.8	167	22	ABG25362	Novel human diagno
987	6	0.8	167	22	AAU18300	Human endocrine po
988	6	0.8	168	19	AA058974	S. pneumoniae deri
989	6	0.8	168	22	ABG21044	Novel human diagno
990	6	0.8	168	22	AA075001	Human colon cancer
991	6	0.8	169	18	AA055432	H. pylori ORF hp4p
992	6	0.8	169	18	AA016321	Murine ARF-p19, a
993	6	0.8	169	21	AA019128	Polypeptide isolat
994	6	0.8	169	21	AA079143	Mouse ARF-p19 cell
995	6	0.8	169	22	AB012215	Human hook2 protei
996	6	0.8	169	23	AA025910	Mouse ARF-p19 prot
997	6	0.8	169	23	AA015373	Mouse Arf protein.
998	6	0.8	172	21	AA013060	Arabidopsis thalia
999	6	0.8	173	22	AAU66992	Propionibacterium
1000	6	0.8	173	22	AAU38850	Novel subtilisin h

PR	07-MAY-1999;	99GB-0010710.
XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.	
PA	Defrenne C, Delmelle C, Ruelle J;	
PI	WPI; 2000-602119/57.	
XX	N-PSDB; AAA75744.	
DR	Novel polypeptides designated BASB 082, 083, 091, 092, and 101 derived	
XX	from meningococcus bacterium useful for producing vaccines against	
PT	infections and in diagnostic assays -	
XX	Claim 3; Page 99-101; 108pp; English.	
PS	The present sequence represents a BASB082 polypeptide. The specification	
XX	describes BASB082, BASB083, BASB091, BASB092, and BASB101 polypeptides.	
CC	The polynucleotides and polypeptides are useful as diagnostic reagents	
CC	and for diagnosing N. meningitidis infection. The polynucleotides may	
CC	be used as hybridisation probe for RNA, cDNA and genomic DNA to isolate	
CC	full-length cDNAs and genomic clones encoding BASB082, BASB083, BASB091,	
CC	BASB092 or BASB101 polypeptides and to isolate cDNA and genomic clones	
CC	of other genes that have a high identity particularly high sequence	
CC	identity to BASB082, BASB083, BASB091, BASB092 or BASB101 genes. The	
CC	vaccine compositions are useful for inducing an immunological response	
CC	in humans. The polynucleotides encoding BASB082, BASB083, BASB091,	
CC	BASB092 or BASB101 polypeptides are useful in gene therapy to induce	
CC	an immunological response. The polypeptides are useful for treating	
CC	upper respiratory tract infection, invasive bacterial diseases, such as	
CC	bacteremia and meningitis.	
XX	SQ Sequence 758 AA;	

Qy	1	MAQTTLPKPIVLSILLINPPLLAQAHETEQSGVLETTVVGKSRPRATSGLLHTSTASDKI	60
Db	1	MAQTTLPKPIVLSILLINPPLLAQAHETEQSGVLETTVVGKSRPRATSGLLHTSTASDKI	60
Qy	61	ISGDTLRQKAVNLGDALDVGPGIHASQYGGASAPVIRGOTGRIKVLNHHGEGDMADF	120
Db	61	ISGDTLRQKAVNLGDALDVGPGIHASQYGGASAPVIRGOTGRIKVLNHHGEGDMADF	120
Qy	121	SPDHAIWDTALTSQVEILRGPVTLTYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRL	180
Db	121	SPDHAIWDTALTSQVEILRGPVTLTYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRL	180
Qy	181	SSGNLEKLTSGGINIGLGNFVLTHTGLYKSGDYAVPRYRNKRLPDSPPRFANGQHRA	240
Db	181	SSGNLEKLTSGGINIGLGNFVLTHTGLYKSGDYAVPRYRNKRLPDSPPRFANGQHRA	240
Qy	241	VLGWRKRFYRTYSDRDQYGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHLLTREDV	300
Db	241	VLGWRKRFYRTYSDRDQYGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHLLTREDV	300
Qy	301	DYDNPGLSCGFHDDDDAHAAHNGKPMWDLNKRKYELRAEWKQFPFGFEALRVHLNRNDY	360
Db	301	DYDNPGLSCGFHDDDDAHAAHNGKPMWDLNKRKYELRAEWKQFPFGFEALRVHLNRNDY	360
Qy	361	HHDEKAGDAVENFNQNTQARIELRHQPIGRKLGSGVQYLGQSSALSATSBAVQKPM	420
Db	361	HHDEKAGDAVENFNQNTQARIELRHQPIGRKLGSGVQYLGQSSALSATSBAVQKPM	420
Qy	421	LLDNKQVHYSFFGVEQANWDFTEGGVRYEKOKASTRYDKALIDRENYKQPLPDLAGH	480
Db	421	LLDNKQVHYSFFGVEQANWDFTEGGVRYEKOKASTRYDKALIDRENYKQPLPDLAGH	480
Qy	481	ROTARSPALSGNMYFFTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFFVGNKHLNKR	540
Db	481	ROTARSPALSGNMYFFTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFFVGNKHLNKR	540

ALIGNMENTS

RESULT 1	
AA018719	
AA018719 standard; Protein; 758 AA.	
AC	AA018719;
XX	
DT	22-JAN-2001 (first entry)
XX	
DE	A Neisseria meningitidis BASB082 polypeptide.
XX	
KW	BASB082; BASB083; BASB091; BASB092; BASB101; infection; vaccine;
KW	gene therapy; upper respiratory tract infection; bacteremia; meningitis;
XX	invasive bacterial disease.
OS	Neisseria meningitidis.
XX	
PN	W020055327-A2.
XX	
PD	21-SEP-2000.
XX	
PF	07-MAR-2000; 2000WO-EP01955.
XX	
PR	12-MAR-1999; 99GB-0005815.
PR	21-APR-1999; 99GB-0009094.
PR	23-APR-1999; 99GB-0009503.
PR	28-APR-1999; 99GB-0009787.

QY 541 SNNIELALGVGDRWOYNLALYRNFGNYIYAQTLDNDRGPKSIEDDSEMKLYVYNQSGA 600
 DB 541 SNNIELALGVGDRWOYNLALYRNFGNYIYAQTLDNDRGPKSIEDDSEMKLYVYNQSGA 600
 QY 601 DFYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQDDQONAPRVP 660
 DB 601 DFYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQDDQONAPRVP 660
 QY 661 AARLGFHLKASLTDRIANLDYVYRFAQNKLARYETRTPGHMLNLGANYRRNTRYGEWN 720
 DB 661 AARLGFHLKASLTDRIANLDYVYRFAQNKLARYETRTPGHMLNLGANYRRNTRYGEWN 720
 QY 721 WYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758
 DB 721 WYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758
 RESULT 2
 AAY74562
 ID AAY74562 standard; Protein; 339 AA.
 XX
 AC AAY74562;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:598.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253324.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 419- ; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 339 AA;
 Query Match 37.7%; Score 286; DB 21; Length 339;
 Best Local Similarity 100.0%; Pred. No. 3.8e-291;
 Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 473 PLFDLGAHQRTARSAFALSGNWYFTPOHKLSTASHOERLPSTOELYAHGKHVATNTFEVG 532
 DB 54 PLFDLGAHQRTARSAFALSGNWYFTPOHKLSTASHOERLPSTOELYAHGKHVATNTFEVG 113
 QY 533 NKHLNKRSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNDRGPKSIEDDSEMKL 592
 DB 114 NKHLNKRSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNDRGPKSIEDDSEMKL 173
 QY 593 VRYNQGADFYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
 DB 174 VRYNQGADFYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 233
 QY 653 DQNAFVPAARLGFHLKASLTDRIANLDYVYRFAQNKLARYETRTPGHMLNLGANYRR 712
 DB 234 DQNAFVPAARLGFHLKASLTDRIANLDYVYRFAQNKLARYETRTPGHMLNLGANYRR 293
 QY 713 NTRYGEWNWYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758
 DB 294 NTRYGEWNWYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 339
 RESULT 3
 AAY74556
 ID AAY74556 standard; Protein; 735 AA.
 XX
 AC AAY74556;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:586.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253318.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2; Page 413; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AA254616 to AA254617 represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides and polypeptides. AA254537 to AA254576 and AA254616 to AA254617 represent PCR primers used in the amplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of and the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to *Neisseria meningitidis* (e.g. meningitis and septicaemia), to detect the presence of *Neisseria meningitidis*, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

XX SQ Sequence 735 AA;

Query Match 37.7%; Score 286; DB 21; Length 735;
Best Local Similarity 100.0%; Pred. No. 8e-291;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPLGAHROTARSAFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVAINTFEVG 532
Db 450 PLPLGAHROTARSAFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVAINTFEVG 509
Qy 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNFRGNYIYAQTLDNGRGPKEIEDDSEMKL 592
Db 510 NKHLKERSNNIELALGYEGDRWQYNLALYRNFRGNYIYAQTLDNGRGPKEIEDDSEMKL 569
Qy 593 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
Db 570 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 629
Qy 653 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 712
Db 630 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 689
Qy 713 NTRYGEWNWYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 690 NTRYGEWNWYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 735

RESULT 4
AAU73011 standard; Protein; 758 AA.
AAU73011;
12-MAR-2002 (first entry)
Neisseria meningitidis virulence protein #101.
Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
infection; Gram-negative bacteria; antimicrobial.
Neisseria meningitidis.
WO200185772-A2.
15-NOV-2001.
08-MAY-2001; 2001WO-GB02003.
08-MAY-2000; 2000GB-0011108.
(MICR-) MICROSCIENCE LTD.
Tang C;
WPI; 2002-066593/09.

DR N-PSDB; AAS97296.
XX New peptide encoded by operon including virulence genes of *Neisseria meningitidis*, useful as vaccine component for treating or preventing meningitis and for identifying antimicrobial drug -
XX Claim 4; Page 412-415; 423pp; English.
XX The invention relates to a peptide (I) encoded by an operon (II) of *Neisseria meningitidis* including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in a Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by *Neisseria* or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAU72911-AAU73014 represent *N. meningitidis* virulence proteins of the invention.

XX SQ Sequence 758 AA;

Query Match 37.7%; Score 286; DB 23; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.2e-291;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPLGAHROTARSAFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVAINTFEVG 532
Db 473 PLPLGAHROTARSAFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVAINTFEVG 532
Qy 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNFRGNYIYAQTLDNGRGPKEIEDDSEMKL 592
Db 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNFRGNYIYAQTLDNGRGPKEIEDDSEMKL 592
Qy 593 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
Db 593 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
Qy 653 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 712
Db 653 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 712
Qy 713 NTRYGEWNWYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 713 NTRYGEWNWYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758

RESULT 5
AAU74565
ID AAU74565 standard; Protein; 758 AA.
XX AAU74565;
AC AAU74565;
XX 21-MAR-2000 (first entry)
DT 21-MAR-2000 (first entry)
XX *Neisseria meningitidis* ORF 149 protein sequence SEQ ID NO:604.
DE *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX *Neisseria meningitidis*.
OS *Neisseria meningitidis*.
XX WO9957280-A2.
FN 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.

```

PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Pi Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Pi Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53327.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics -
XX
XX Claim 2; Page 422-423; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 758 AA;
XX
XX Query Match 25.3%; Score 192; DB 21; Length 758;
XX Best Local Similarity 100.0%; Pred. No. 4.4e-192;
XX Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 38 VVGKSRPRATSGLLHTSTASDKIISGTLRQKAVNLGDLGVPGIHASQYGGASAPVI 97
XX Db ///////////////////////////////////////////////////////////////////
XX 38 VVGKSRPRATSGLLHTSTASDKIISGTLRQKAVNLGDLGVPGIHASQYGGASAPVI 97
XX
XX QY 98 RGGTGRIKVLNHHGETGMDADSPDHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVD 157
XX Db ///////////////////////////////////////////////////////////////////
XX 98 RGGTGRIKVLNHHGETGMDADSPDHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVD 157
XX
XX QY 158 VADGKIPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAV 217
XX Db ///////////////////////////////////////////////////////////////////
XX 158 VADGKIPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAV 217
XX
XX QY 218 PRYNLRKLDPDS 229
XX Db ///////////////////////////////////////////////////////////////////
XX 218 PRYNLRKLDPDS 229
XX
XX RESULT 6
XX AAZ74563
XX ID AAZ74563 standard; Protein; 339 AA.
XX
XX AC AAZ74563;
XX
XX 21-MAR-2000 (first entry)
XX
XX DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:600.
XX
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.
XX
XX OS Neisseria meningitidis.
XX

```

```

PN WO9957280-A2.
XX
XX PD 11-NOV-1999.
XX
XX PF 30-APR-1999; 99WO-US09346.
XX
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53325.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics -
XX
XX Claim 2; Page 420; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 339 AA;
XX
XX Query Match 23.1%; Score 175; DB 21; Length 339;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-174;
XX Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 473 PLFDLGAHROTARSPALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 532
XX Db ///////////////////////////////////////////////////////////////////
XX 54 PLFDLGAHROTARSPALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 113
XX
XX QY 533 NKHLNERSNNIELALGYEGDRWQYNLALYRNFPGNYIYAQTLLNDGRGPKSIEDDSEMKL 592
XX Db ///////////////////////////////////////////////////////////////////
XX 114 NKHLNERSNNIELALGYEGDRWQYNLALYRNFPGNYIYAQTLLNDGRGPKSIEDDSEMKL 173
XX
XX QY 593 VRYNQSADFYGAEGEITYFKPTPRYRIGVSGDVYVRGLKNLPSLPGRDAYGNRP 647
XX Db ///////////////////////////////////////////////////////////////////
XX 174 VRYNQSADFYGAEGEITYFKPTPRYRIGVSGDVYVRGLKNLPSLPGRDAYGNRP 228
XX
XX RESULT 7
XX AAZ74557
XX ID AAZ74557 standard; Protein; 764 AA.
XX
XX AC AAZ74557;
XX
XX 21-MAR-2000 (first entry)
XX
XX DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:588.
XX

```

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

OS Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

XX Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AAZ53319.

XX Novel Neisserial polypeptides predicted to be useful antigens for

XX vaccines and diagnostics -

XX Claim 2; Page 414-415; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

SQ Sequence 764 AA;

Query Match 23.1%; Score 175; DB 21; Length 764;

Best Local Similarity 100.0%; Pred. No. 3.1e-174;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 PLPDGAHQRTARFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVATNTFEVG 532

DB 479 PLPDGAHQRTARFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVATNTFEVG 538

QY 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFNGNYIYAQTLDNDRGPKSIEDDSEMKL 592

DB 539 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFNGNYIYAQTLDNDRGPKSIEDDSEMKL 598

QY 593 VRYNQSGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 647

DB 599 VRYNQSGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 653

RESULT 8

ABB78068

ID ABB78068 standard; Protein; 764 AA.

XX

AC ABB78068;

XX 05-NOV-2002 (first entry)

XX Amino acid sequence of p88 polypeptide.

XX p177; p88; p64; p55; p46; vaccine; gonorrhea.

XX Neisseria gonorrhoeae.

XX WO200260936-A2.

XX 08-AUG-2002.

XX 31-JAN-2002; 2002WO-US02881.

XX 31-JAN-2001; 2001US-266070P.

XX 06-AUG-2001; 2001US-310356P.

XX 23-OCT-2001; 2001US-34452P.

XX (IOWA) UNIV IOWA RES FOUND.

XX (REGC) UNIV CALIFORNIA.

XX (APIC/) APICELLA M. A.

XX (EDWA/) EDWARDS J. L.

XX (GIES/) GIBSON B. W.

XX (SCHE/) SCHEFFLER K.

XX (BROW/) BROWN E.

XX Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;

XX WPI; 2002-619227/66.

XX N-PSDB; ABQ78299.

XX New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria

XX gonorrhoeae, useful for preventing, or protecting a female patient

XX against, N. gonorrhoeae colonization or infection -

XX Claim 7; Page 115-117; 130pp; English.

XX The present sequence represents a p88 polypeptide. The specification

XX describes p177, p88, p64, p55 and p46 polypeptides from Neisseria

XX gonorrhoeae. The polypeptides are useful as vaccines, for preventing,

XX or protecting a female patient against, N. gonorrhoeae colonization or

XX infection. Such immunisation can prevent gonorrhea in women.

XX Sequence 764 AA;

Query Match 23.1%; Score 175; DB 23; Length 764;

Best Local Similarity 100.0%; Pred. No. 3.1e-174;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 PLPDGAHQRTARFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVATNTFEVG 532

DB 479 PLPDGAHQRTARFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVATNTFEVG 538

QY 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFNGNYIYAQTLDNDRGPKSIEDDSEMKL 592

DB 539 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFNGNYIYAQTLDNDRGPKSIEDDSEMKL 598

QY 593 VRYNQSGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 647

DB 599 VRYNQSGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 653

RESULT 9

ABP80923

ID ABP80923 standard; Protein; 728 AA.

XX

AC ABP80923;

XX

DT 07-MAR-2003 (first entry)

XX

DE N. gonorrhoeae amino acid sequence SEQ ID 8376.

XX Antibacterial; infection; vaccine; gene therapy.
 KW Neisseria gonorrhoeae.
 OS WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizza M, Masignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX N-PSDB; ABZ41893.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 XX Disclosure; Page 799; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX SQ Sequence 728 AA;
 Query Match 18.2%; Score 138; DB 24; Length 728;
 Best Local Similarity 100.0%; Pred. No. 2.2e-135;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGGASAPVI 97
 DB 8 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGGASAPVI 67
 QY 98 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPVTLTYSSGNVAGLVD 157
 DB 68 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPVTLTYSSGNVAGLVD 127
 QY 158 VADGKIPEKMPENGVSIGE 175
 DB 128 VADGKIPEKMPENGVSIGE 145
 RESULT 10
 AAY74564
 ID AAY74564 standard; Protein; 758 AA.
 XX AAY74564;
 XX 21-MAR-2000 (first entry)
 XX Neisseria gonorrhoeae ORF 149 protein sequence SEQ ID NO:602.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX Neisseria gonorrhoeae.
 OS WO9957280-A2.
 XX 11-NOV-1999.
 XX

PF 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-PSDB; AAZ53326.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 2; Page 421-422; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ54253 to AAZ54941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX SQ Sequence 758 AA;
 Query Match 18.2%; Score 138; DB 21; Length 758;
 Best Local Similarity 100.0%; Pred. No. 2.3e-135;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGGASAPVI 97
 DB 38 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGGASAPVI 97
 QY 98 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPVTLTYSSGNVAGLVD 157
 DB 98 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPVTLTYSSGNVAGLVD 157
 QY 158 VADGKIPEKMPENGVSIGE 175
 DB 158 VADGKIPEKMPENGVSIGE 175
 RESULT 11
 AAY74566
 ID AAY74566 standard; Protein; 758 AA.
 XX AAY74566;
 XX 21-MAR-2000 (first entry)
 XX Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:606.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX

```

OS Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53328.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 424-425; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitis and N. gonorrhea polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX Sequence 758 AA;
XX Query Match 17.9%; Score 136; DB 21; Length 758;
XX Best Local Similarity 100.0%; Pred. No. 2.8e-133;
XX Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 473 PLPDLGAHROTARSPALSGNWFYTPQHKLSLTASHQERLPSTQELIYAHGKHVATNTFVG 532
Db 473 PLPDLGAHROTARSPALSGNWFYTPQHKLSLTASHQERLPSTQELIYAHGKHVATNTFVG 532
QY 533 NKHLNKRSNNIELALGYEGDRWQYNLALYRNRFNGNIVYAOTLNDGRGPKSTIEDSENKL 592
Db 533 NKHLNKRSNNIELALGYEGDRWQYNLALYRNRFNGNIVYAOTLNDGRGPKSTIEDSENKL 592
QY 593 VRYNQSGADFYGAEGE 608
Db 593 VRYNQSGADFYGAEGE 608
RESULT 12
AAZ74555
ID AAZ74555 standard; Protein; 174 AA.
XX AC AAZ74555;
XX AC AAZ74555;
DT 21-MAR-2000 (first entry)
XX

```

```

DE Neisseria gonorrhea ORF 147 protein sequence SEQ ID NO:584.
XX Neisseria meningitidis; Neisseria gonorrhea; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibacterial; gene therapy.
XX OS Neisseria gonorrhea.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53317.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 412; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitis and N. gonorrhea polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX Sequence 174 AA;
XX Query Match 15.4%; Score 117; DB 21; Length 174;
XX Best Local Similarity 100.0%; Pred. No. 6.3e-114;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 VVGKSRPRATSGLLHTSTASDKIISGDTLRKAVNLGDLGVPGIHASQYGGGASAFVI 97
Db 44 VVGKSRPRATSGLLHTSTASDKIISGDTLRKAVNLGDLGVPGIHASQYGGGASAFVI 103
QY 98 RGQTGRRIKVLNHHGETGDMADFPDHALIWDVTALSSQVETLRGPVTLIYSSGNVAG 154
Db 104 RGQTGRRIKVLNHHGETGDMADFPDHALIWDVTALSSQVETLRGPVTLIYSSGNVAG 160
RESULT 13
AAZ74561
ID AAZ74561 standard; Protein; 339 AA.
XX AC AAZ74561;
XX AC AAZ74561;

```

```
DT 21-MAR-2000 (first entry)
XX Neisseria gonorrhoeae ORF 149 protein sequence SEQ ID NO:596.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX Neisseria gonorrhoeae.
OS
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHTR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
XX N-PSDB; AAZ53323.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
XX Claim 2; Page 418; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 339 AA;
SQ
Query Match 10.6%; Score 80; DB 21; Length 339;
Best Local Similarity 100.0%; Pred. No. 8.6e-75;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 562 YNRNFGNTIYAQTLDNGRGPKEIEDSEMKLVRYNQSGADFYGAGEIYFKPTPRYRIGV 621
Db 143 YNRNFGNTIYAQTLDNGRGPKEIEDSEMKLVRYNQSGADFYGAGEIYFKPTPRYRIGV 202
QY 622 SGDYVRGRILKNLPSLPGRED 641
Db 203 SGDYVRGRILKNLPSLPGRED 222
RESULT 14
AAG15254
ID AAG15254 standard; Protein; 144 AA.
XX
```

```
AC AAG15254;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 15434.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
```


PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
|||||
Db 9 ALSATSEAV 17

RESULT 15

AG15253
ID AAG15253 standard; Protein; 145 AA.

XX
AC AAG15253;

XX
DT 17-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 15433.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;

```
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 23-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 23-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
```

PR	07-SEP-1999;	99US-0152363.	PD	06-SEP-2000.	
PR	10-SEP-1999;	99US-0153070.	XX		
PR	13-SEP-1999;	99US-0153758.	PF	25-FEB-2000;	2000EP-0301439.
PR	15-SEP-1999;	99US-0154018.	XX		
PR	16-SEP-1999;	99US-0154039.	XX	25-FEB-1999;	99US-0121825.
PR	20-SEP-1999;	99US-0154779.	PR	05-MAR-1999;	99US-0123180.
PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999;	99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0127462.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159331.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159637.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132486.
PR	18-OCT-1999;	99US-0159584.	PR	06-MAY-1999;	99US-0132487.
PR	21-OCT-1999;	99US-0160741.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160767.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR	14-MAY-1999;	99US-0134370.
PR	22-OCT-1999;	99US-0160980.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160981.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0161404.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161406.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161359.	PR	25-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161992.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162142.	PR	29-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142055.

Query Match 1.2%; Score 9; DB 21; Length 145;
Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 408 ALSATSEAV 416

Db 10 ALSATSEAV 18

RESULT 16

AAG15252

ID AAG15252 standard; Protein; 151 AA.

XX

AC AAG15252;

XX

DT 17-OCT-2000 (first entry)

DE

XX Arabidopsis thaliana protein fragment SEQ ID NO: 15432.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

```
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147240.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148585.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Gaps 0;
Indels 0; Gaps 0;

Qy 408 ALSATSEAV 416
Db 16 ALSATSEAV 24

RESULT 17
AAG07631
ID AAG07631 standard; Protein; 323 AA.
XX AAG07631;
AC AAG07631;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 4857.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
```

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR	07-OCT-1999;	99US-0158023.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134370.
PR	22-OCT-1999;	99US-0160980.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160981.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	20-MAY-1999;	99US-0135124.
PR	25-OCT-1999;	99US-0161404.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161406.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161359.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161922.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137724.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUN-1999;	99US-0138094.
			PR	10-JUN-1999;	99US-0138540.
			PR	10-JUN-1999;	99US-0138847.
			PR	14-JUN-1999;	99US-0139119.
			PR	16-JUN-1999;	99US-0139452.
			PR	16-JUN-1999;	99US-0139453.
			PR	17-JUN-1999;	99US-0139492.
			PR	18-JUN-1999;	99US-0139454.
			PR	18-JUN-1999;	99US-0139455.
			PR	18-JUN-1999;	99US-0139456.
			PR	18-JUN-1999;	99US-0139457.
			PR	18-JUN-1999;	99US-0139458.
			PR	18-JUN-1999;	99US-0139459.
			PR	18-JUN-1999;	99US-0139460.
			PR	18-JUN-1999;	99US-0139461.
			PR	18-JUN-1999;	99US-0139462.
			PR	18-JUN-1999;	99US-0139463.
			PR	18-JUN-1999;	99US-0139750.
			PR	18-JUN-1999;	99US-0139763.
			PR	21-JUN-1999;	99US-0139817.
			PR	22-JUN-1999;	99US-0139899.
			PR	23-JUN-1999;	99US-0140353.
			PR	23-JUN-1999;	99US-0140354.
			PR	24-JUN-1999;	99US-0140695.
			PR	28-JUN-1999;	99US-0140823.
			PR	29-JUN-1999;	99US-0140981.
			PR	30-JUN-1999;	99US-0141287.
			PR	01-JUL-1999;	99US-0141842.
			PR	01-JUL-1999;	99US-0142154.
			PR	02-JUL-1999;	99US-0142055.
			PR	06-JUL-1999;	99US-0142390.
			PR	08-JUL-1999;	99US-0142803.
			PR	09-JUL-1999;	99US-0142920.
			PR	12-JUL-1999;	99US-0142977.
			PR	13-JUL-1999;	99US-0143542.
			PR	14-JUL-1999;	99US-0143624.
			PR	15-JUL-1999;	99US-0144005.
			PR	16-JUL-1999;	99US-0144085.
			PR	16-JUL-1999;	99US-0144086.
			PR	19-JUL-1999;	99US-0144325.
			PR	19-JUL-1999;	99US-0144331.
			PR	19-JUL-1999;	99US-0144332.
			PR	19-JUL-1999;	99US-0144333.
			PR	19-JUL-1999;	99US-0144334.

Query Match 1.2%; Score 9; DB 21; Length 323;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 408 ALSATSEAV 416
Db 9 ALSATSEAV 17

RESULT 18
AAG36807
ID AAG36807 standard; Protein; 323 AA.
XX AC AAG36807;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 45159.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EF1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.

PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145112.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150586.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151920.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160815.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 323;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 408 ALSATSEAV 416
 Db 9 ALSATSEAV 17

RESULT 19
 ABB93042
 ID ABB93042 standard; Protein; 323 AA.
 XX AC ABB93042;
 XX 31-MAY-2002 (first entry)
 DT Herbicidally active polypeptide SEQ ID NO 2253.
 DE Herbicidally active polypeptide
 DE Herbicid; plant; agriculture; herbicide.
 KW Arabidopsis thaliana.
 XX OS W0200210210-A2.
 XX PD 07-FEB-2002.
 XX PF 28-AUG-2001; 2001WO-EP09892.
 XX PR 28-AUG-2001; 2001WO-EP09892.
 XX PA (FARB) BAYER AG.
 XX Tietjen K, Weidler M;
 XX WPI; 2002-269010/31.
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX Claim 5; SEQ ID NO 2253; 261pp + Sequence Listing; English.
 XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX

SQ Sequence 323 AA;

Query Match 1.2%; Score 9; DB 23; Length 323;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416

Db 9 ALSATSEAV 17
|||||||

RESULT 20

AAG07630
ID AAG07630 standard; Protein; 324 AA.

XX AAG07630;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 4856.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; Genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 03-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127482.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 04-MAY-1999; 99US-0132048.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 20-MAY-1999; 99US-0134941.

XX 21-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 25-MAY-1999; 99US-0135629.

XX 27-MAY-1999; 99US-0136021.

XX 99US-0136392.

XX

XX

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140821.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.

```
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 324;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
Db 10 ALSATSEAV 18

RESULT 21
AAG36806
ID AAG36806 standard; Protein; 324 AA.
XX
AC AAG36806;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45158.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134216.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
```

PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 24-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140591.
 PR 29-JUN-1999; 99US-0141287.
 PR 30-JUN-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142190.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142820.
 PR 09-JUL-1999; 99US-0142977.
 PR 12-JUL-1999; 99US-0143542.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145313.
 PR 27-JUL-1999; 99US-0145318.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149829.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 26-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 25-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161933.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 324;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 408 ALSATSEAV 416
 |||||

Db 10 ALSATSEAV 18

RESULT 22
AAG07629
ID AAG07629 standard; Protein; 330 AA.
XX AAG07629;
AC AAG07629;
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 4855.
DE Arabidopsis thaliana
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136399.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137503.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139499.
PR 18-JUN-1999; 99US-0139499.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147152.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR	18-AUG-1999;	99US-0149426.	AC	AAG36805;
PR	20-AUG-1999;	99US-0149722.	XX	
PR	20-AUG-1999;	99US-0149723.	DT	18-OCT-2000 (first entry)
PR	20-AUG-1999;	99US-0149929.	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 45157.
PR	23-AUG-1999;	99US-0149902.	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 45157.
PR	23-AUG-1999;	99US-0149930.	XX	Protein identification; signal transduction pathway; metabolic pathway;
PR	25-AUG-1999;	99US-0150566.	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
PR	25-AUG-1999;	99US-0150884.	KW	termination sequence.
PR	26-AUG-1999;	99US-0151065.	XX	
PR	27-AUG-1999;	99US-0151066.	OS	Arabidopsis thaliana.
PR	27-AUG-1999;	99US-0151080.	XX	
PR	27-AUG-1999;	99US-0151303.	PN	EPI033405-A2.
PR	30-AUG-1999;	99US-0151438.	XX	06-SEP-2000.
PR	31-AUG-1999;	99US-0151930.	PD	
PR	01-SEP-1999;	99US-0152363.	XX	25-FEB-2000; 2000EP-0301439.
PR	07-SEP-1999;	99US-0153070.	XX	
PR	10-SEP-1999;	99US-0153758.	XX	25-FEB-1999; 99US-0121825.
PR	13-SEP-1999;	99US-0154018.	PR	05-MAR-1999; 99US-0123180.
PR	15-SEP-1999;	99US-0154039.	PR	09-MAR-1999; 99US-0123548.
PR	16-SEP-1999;	99US-0154779.	PR	23-MAR-1999; 99US-0125788.
PR	20-SEP-1999;	99US-0155139.	PR	25-MAR-1999; 99US-0126264.
PR	22-SEP-1999;	99US-0155486.	PR	29-MAR-1999; 99US-0126785.
PR	23-SEP-1999;	99US-0155659.	PR	01-APR-1999; 99US-0127452.
PR	24-SEP-1999;	99US-0156458.	PR	06-APR-1999; 99US-0128234.
PR	28-SEP-1999;	99US-0156596.	PR	08-APR-1999; 99US-0128714.
PR	04-OCT-1999;	99US-0157117.	PR	16-APR-1999; 99US-0129845.
PR	05-OCT-1999;	99US-0157753.	PR	19-APR-1999; 99US-0130077.
PR	06-OCT-1999;	99US-0157865.	PR	21-APR-1999; 99US-0130449.
PR	07-OCT-1999;	99US-0158029.	PR	23-APR-1999; 99US-0130510.
PR	08-OCT-1999;	99US-0158232.	PR	28-APR-1999; 99US-0130891.
PR	12-OCT-1999;	99US-0158369.	PR	30-APR-1999; 99US-0131449.
PR	13-OCT-1999;	99US-0159293.	PR	30-APR-1999; 99US-0132048.
PR	13-OCT-1999;	99US-0159294.	PR	04-MAY-1999; 99US-0132407.
PR	14-OCT-1999;	99US-0159329.	PR	05-MAY-1999; 99US-0132484.
PR	14-OCT-1999;	99US-0159330.	PR	06-MAY-1999; 99US-0132485.
PR	14-OCT-1999;	99US-0159331.	PR	06-MAY-1999; 99US-0132486.
PR	14-OCT-1999;	99US-0159637.	PR	07-MAY-1999; 99US-0132863.
PR	14-OCT-1999;	99US-0159638.	PR	11-MAY-1999; 99US-0134256.
PR	18-OCT-1999;	99US-0159584.	PR	14-MAY-1999; 99US-0134218.
PR	21-OCT-1999;	99US-0160741.	PR	14-MAY-1999; 99US-0134219.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999; 99US-0134221.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999; 99US-0134370.
PR	21-OCT-1999;	99US-0160770.	PR	18-MAY-1999; 99US-0134768.
PR	21-OCT-1999;	99US-0160814.	PR	19-MAY-1999; 99US-0134941.
PR	21-OCT-1999;	99US-0160815.	PR	20-MAY-1999; 99US-0135124.
PR	22-OCT-1999;	99US-0160980.	PR	21-MAY-1999; 99US-0135353.
PR	22-OCT-1999;	99US-0160981.	PR	24-MAY-1999; 99US-0135629.
PR	22-OCT-1999;	99US-0160989.	PR	25-MAY-1999; 99US-0136021.
PR	25-OCT-1999;	99US-0161404.	PR	27-MAY-1999; 99US-0136392.
PR	25-OCT-1999;	99US-0161405.	PR	28-MAY-1999; 99US-0136782.
PR	26-OCT-1999;	99US-0161359.	PR	01-JUN-1999; 99US-0137222.
PR	26-OCT-1999;	99US-0161360.	PR	03-JUN-1999; 99US-0137528.
PR	26-OCT-1999;	99US-0161361.	PR	04-JUN-1999; 99US-0137502.
PR	28-OCT-1999;	99US-0161920.	PR	07-JUN-1999; 99US-0137724.
PR	28-OCT-1999;	99US-0161932.	PR	08-JUN-1999; 99US-0138094.
PR	28-OCT-1999;	99US-0161993.	PR	10-JUN-1999; 99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999; 99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999; 99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999; 99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999; 99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999; 99US-0139492.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999; 99US-0139462.

Query Match 1.24; Score 9; DB 21; Length 330;
 Best Local Similarity 100.0%; Pred. No. 3.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
 |||||
 Db 16 ALSATSEAV 24

RESULT 23
 AAG36805
 ID AAG36805 standard; Protein; 330 AA.
 XX

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159323.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
|||
Db 16 ALSATSEAV 24

RESULT 24
ABR41659
ID ABR41659 standard; Protein; 1240 AA.
XX
AC ABR41659;
XX
DT 02-JUN-2003 (first entry)
XX
DE Human DITHP cytoskeletal protein.
XX
KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;

KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW cytoskeletal protein.

OS Homo sapiens.

XX WO200297031-A2.

FN 05-DEC-2002.

PD 27-MAR-2002; 2002WO-US10056.

XX 28-MAR-2001; 2001US-279619P.

PR 29-MAR-2001; 2001US-280067P.

PR 29-MAR-2001; 2001US-280068P.

PR 16-MAY-2001; 2001US-291280P.

PR 17-MAY-2001; 2001US-291289P.

PR 17-MAY-2001; 2001US-291849P.

PR 19-JUN-2001; 2001US-299428P.

PR 20-JUN-2001; 2001US-299776P.

PR 20-JUN-2001; 2001US-300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;

PI Doughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Feralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-129518/12.

DR N-PSDB; ACC46596.

DR Novel human diagnostic and therapeutic polypeptide useful for

XX identifying test compound which specifically binds to a polypeptide

PT encoded by human diagnostic and therapeutic polynucleotide, and to

PT induce antibodies -

XX Claim 27; SEQ ID No 1194; 591pp; English.

PS The invention relates to novel human diagnostic and therapeutic

CC polynucleotides designated dithp (ACC46080-ACC46749) and to their

CC encoded proteins (DITHP; ABK4136-ABK41812). The invention also relates

CC to polynucleotide sequences at least 90% identical to the dithp cDNA

CC sequences of the invention; recombinant vectors, host cells and

CC transgenic organisms comprising a dithp nucleic acid sequence; the

CC recombinant production of DITHP proteins; antibodies specific for DITHP

CC proteins; microarrays comprising dithp nucleic acid sequences; methods

CC of detecting dithp nucleotide and protein sequences; methods of screening

CC for compounds which specifically bind a DITHP protein; and methods of

CC assessing the toxicity of test compounds using a dithp hybridisation

CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the

CC diagnosis of a wide variety of conditions including cancer and other cell

CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,

CC viral, fungal or parasitic infections; hormonal disorders; metabolic

CC disorders; neurological disorders; gastrointestinal disorders; transport

CC disorders; and connective tissue disorders. They may also be used to

CC screen for modulators of protein activity or gene expression. DITHP

CC proteins can additionally be used in analysis of the proteome of a tissue

CC or cell type and to induce antibodies. The dithp nucleic acids are

CC additionally useful in somatic or germline gene therapy of the disorders

CC mentioned above, as a source of antisense sequences, as a source of

CC probes and primers, in genotyping and identification of individuals, in

CC the generation of transgenic animal models of human disease or knock in

CC humanised animals, in toxicological testing, and in transcript imaging.

CC The present sequence represents a DITHP protein which is a cytoskeletal

CC protein.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1240 AA;

SQ Query Match 1.2%; Score 9; DB 24; Length 1240;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSG 213

Db 939 TEGLYRKSG 947

RESULT 25

ABP58986

ID ABP58986 standard; Protein; 94 AA.

XX AC ABP58986;

DT 14-APR-2003 (first entry)

XX Human Mch2 protein 10.34.

Human; Mch2 protein 10.34; recombinant production; gene therapy;

neuropathy; tumour; cancer; embryonic development disorder;

growth disorder; immune disorder; inflammatory condition; cytostatic;

antiinflammatory; immunomodulator.

OS Homo sapiens.

XX CN1361168-A.

PD 31-JUL-2002.

XX 26-DEC-2000; 2000CN-0136339.

XX 26-DEC-2000; 2000CN-0136339.

XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI; 2003-000242/01.

XX N-PSDB; ABZ70869.

XX New polypeptide human Mch2 protein 10.34 and polynucleotides encoding

XX this polypeptide -

XX Claim 1; Page 28 (Disclosure); 34pp; Chinese.

XX The invention relates to human Mch2 protein 10.34 (ABP58986) and nucleic

XX acids encoding it (ABZ70869). The protein has a molecular weight of

XX 10.34 kD. The invention also relates to a method for the recombinant

XX production of the protein, an antagonist of the protein, and the use of

XX the protein, gene and antagonist in therapeutic applications. Mch2

XX protein 10.34 can be used in the treatment of a variety of diseases such

XX as neuropathy, tumours, embryonic development disorders, growth

XX disorders, immune disorders and inflammatory conditions. The present

XX sequence represents human Mch2 protein 10.34.

XX Sequence 94 AA;

QY 96 VIRGQTGR 103

Db 37 VIRGQTGR 44

RESULT 26

ABB47432


```

ID  ABB47432 standard; Protein; 99 AA.
XX  ABB47432;
XX
XX  05-FEB-2002 (first entry)
XX
XX  Listeria monocytogenes protein #136.
DE
XX  Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW  vitamin B12; bacterial infection; disease.
XX
XX  Listeria monocytogenes.
XX
XX  W0200177335-A2.
XX
XX  18-OCT-2001.
XX
XX  11-APR-2001; 2001WO-FR011118.
XX
XX  11-APR-2000; 2000FR-0004629.
PR
XX  (INSP ) INST PASTEUR.
PA
XX
XX  Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;
PI  Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI  Daniels J, Goebel W, Kraft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI  Dominguez-Bernal G, Garrido-Garcia P, Tlerrrez-Martinez A, Amend A;
PI  Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant N;
PI  Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI  Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI  Rose M, Voss H;
XX
XX  WPI; 2002-010914/01.
DR
XX
XX  Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT  and prevention of Listeria and related bacterial infections, and
PT  related polypeptides -
XX
XX  Claim 6; SEQ ID No 137; 132pp; French.
PS
XX
XX  The present invention relates to the genome sequence of Listeria
CC  monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC  it are useful for selecting probes and primers for detecting genes in L.
CC  monocytogenes and related organisms, and for studying genetic
CC  polymorphisms and other genomes. The present sequence is a protein
CC  encoded by the genome sequence of the present invention. Proteins
CC  expressed from the genome sequence are useful for raising specific
CC  antibodies, identification of L. monocytogenes and related organisms, and
CC  for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC  B12. The genome sequence and proteins encoded by it are also useful for
CC  selecting compounds that regulate gene expression and cell replication
CC  and modulate L. monocytogenes-related diseases. In addition, the genome
CC  sequence and proteins encoded by it are useful in pharmaceutical and
CC  vaccines compositions for the treatment or prevention of infections by L.
CC  monocytogenes and related organisms.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published\_pct\_sequences.
XX
XX  Sequence 99 AA;
SQ
    Query Match 1.1%; Score 8; DB 23; Length 99;
    Best Local Similarity 100.0%; Pred. No. 11;
    Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 AQTTLKPI 9
    Db 46 AQTTLKPI 53
    |||||
    |||||

RESULT 27
ABP32719
ID  ABP32719 standard; Protein; 115 AA.

```

```

XX  ABP32719;
AC
XX
XX  08-JUL-2002 (first entry)
DT
XX
XX  Human ORF1692 protein, SEQ ID NO:3384.
DE
XX
XX  Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW  disease monitoring; cytokine; cell proliferation; cell differentiation;
KW  immune modulation; haematopoiesis regulation; tissue growth;
KW  angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW  thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW  behaviour; cancer; proliferative disorder; neurological disorder;
KW  cardiovascular disease; immune system disorder; organ transplantation;
KW  tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW  hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW  vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW  neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW  cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW  dermatological; analgesic; virucide; antibacterial; fungicide.
XX
XX  Homo sapiens.
OS
XX
XX  W0200190366-A2.
XX
XX  29-NOV-2001.
PD
XX
XX  24-MAY-2001; 2001WO-US17076.
PF
XX
XX  24-MAY-2000; 2000US-206690P.
PR
XX
XX  (CURA-) CURAGEN CORP.
PA
XX
XX  Leach MD, Shimkets RA;
PI
XX
XX  WPI; 2002-106200/14.
DR
XX
XX  N-PSDB; ABN76745.
DR
XX
XX  Novel human polypeptides and polynucleotides useful for diagnosing,
PT  preventing and treating cardiovascular disease, neurodegenerative,
PT  hyperproliferative disorders and disorders related to organ
PT  transplantation -
XX
XX  Claim 10; Page 1105; 2508pp; English.
PS
XX
XX  Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC  designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC  ABN79587 represent cDNAs encoding them. The invention also encompasses
CC  polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC  referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC  the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC  polynucleotides, the recombinant production of ORFX proteins, antibodies
CC  specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC  polypeptides, methods of screening for modulators of ORFX expression or
CC  activity, and methods of screening individuals for a predisposition to an
CC  ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC  range of biological activities, such as cytokine, cell proliferation,
CC  cell differentiation, immune modulation, haematopoiesis regulation,
CC  tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC  chemokinetic activity, haemostatic activity, thrombolytic activity,
CC  receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC  and antiinfective activity, and may also be involved in the determination
CC  of bodily characteristics, fertility and behaviour. ORFX proteins,
CC  nucleic acids and antibodies may be used in the treatment of cancers,
CC  other proliferative disorders such as psoriasis and benign tumours,
CC  neurological disorders such as epilepsy and Alzheimer's disease,
CC  cardiovascular diseases, immune system disorders, disorders related to
CC  organ transplantation, disorders of tissue growth and regeneration,
CC  diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC  storage disease, and infectious diseases caused by viral, bacterial,
CC  fungal and other pathogens. ORFX nucleic acids may also be used as a
CC  source of primers and probes, in the detection of ORFX genomic sequences
CC  or transcripts, in the identification and cloning of homologous

```

CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 XX
 SQ Sequence 115 AA;

Query Match 1.1%; Score 8; DB 23; Length 115;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYRKSG 213
 DB 96 EGLYRKSG 103

RESULT 28
 ABB57898
 ID ABB57898 standard; Protein; 268 AA.
 XX
 AC ABB57898;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 486.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEXE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL02001.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 486; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 268 AA;

Query Match 1.1%; Score 8; DB 22; Length 268;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 RSNNIELA 547
 DB 67 RSNNIELA 74
 ID AAU69473 standard; Protein; 313 AA.
 XX
 AC AAU69473;
 XX
 DT 29-JAN-2002 (first entry)
 DE Human purified secretory polypeptide #42.
 XX
 KW Human; purified secretory polypeptide; cell proliferative disorder;
 KW cancer; immune system disorder; neurological disorder; mental disorder;
 KW motor neuron disorder; demyelinating disorder; neuromuscular disorder;
 KW central nervous system disorder; enzyme linked immunosorbent assay;
 KW ELISA; gene therapy.
 XX
 OS Homo sapiens.
 PN WO200162918-A2.
 PD 30-AUG-2001.
 XX
 PF 01-FEB-2001; 2001WO-US03465.
 XX
 PR 24-FEB-2000; 2000US-185215P.
 PR 16-MAY-2000; 2000US-185216P.
 PR 17-MAY-2000; 2000US-205232P.
 PR 17-MAY-2000; 2000US-205286P.
 PR 17-MAY-2000; 2000US-205287P.
 PR 17-MAY-2000; 2000US-205323P.
 PR 17-MAY-2000; 2000US-205324P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'sa SA, Amehy S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HJ, Hodgson DM, Lincoln SE;
 XX
 DR WPI; 2001-648217/74.
 XX
 PT Nucleic acids encoding secretory polypeptides, useful in genetic
 PT diagnosis and therapy -
 XX
 PS Disclosure; Page 214-215; 237pp; English.
 XX
 CC Sequences AAU69432-AAU69511 represent purified secretory polypeptides of
 CC the invention. The polypeptides and their associated polynucleotides can
 CC be used in the treatment, prevention and diagnosis of diseases associated
 CC with inappropriate secretory protein expression. These diseases include
 CC cell proliferative disorders such as atherosclerosis and psoriasis,
 CC cancers such as leukaemia and melanoma, immune system disorders such as
 CC asthma and diabetes mellitus, neurological disorders such as epilepsy and
 CC Parkinson's disease, mental disorders such as schizophrenia and seasonal
 CC affective disorder (SAD), motor neuron disorders such as amyotrophic
 CC lateral sclerosis, demyelinating disorders such as multiple sclerosis,
 CC central nervous system disorders such as mental retardation and
 CC neurofibromatosis and neuromuscular disorders such as cerebral palsy and
 CC muscular dystrophy. Target polynucleotides in a sample can be detected by
 CC hybridising the sample with a probe sequence complementary to the target
 CC polynucleotide, under conditions in which a hybridisation complex is
 CC formed, and detecting the presence or absence of the complex. The
 CC polypeptides may also be used as antigens in the production of antibodies
 CC against secretory proteins and in assays to identify modulators of
 CC protein expression and activity. The antibodies may also be used as
 CC diagnostic agents for detecting the presence of the sequences of the

DE *Mycobacterium vaccae* formate dehydrogenase C6S/C146S/C256S mutant, XX

XX	Formate dehydrogenase; alcohol production; ketone;
KW	mutant; muten.
XX	
OS	Mycobacterium vaccae.
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 6 /note= "Wild type Cys substituted by Ser"
FT	Misc-difference 146 /note= "Wild type Cys substituted by Ser"
FT	Misc-difference 146 /note= "Wild type Cys substituted by Ser"
FT	Misc-difference 256 /note= "Wild type Cys substituted by Ser"
FT	Misc-difference 256 /note= "Wild type Cys substituted by Ser"
XX	
PN	EP1211316-A1.
XX	
PD	05-JUN-2002.
XX	
PF	27-NOV-2001; 2001EP-0128170.
XX	
PR	29-NOV-2000; 2000JP-0363894.
PR	24-AUG-2001; 2001JP-0254631.
XX	
PA	(DAIL) DAICEL CHEM IND LTD.
XX	
PI	Mitsuhashi K, Yamamoto H, Kimoto N;
XX	
DR	WPI; 2002-464925/50.
XX	
PT	New mutant forms of formate dehydrogenase derived from Myc
PT	vaccae have enhanced activity in the presence of organic
PT	are useful for producing alcohols from ketone raw materia
XX	
PS	Claim 8; Page -; 42pp; English.
XX	
CC	This invention relates to a mutant form of the Mycobacter
CC	formate dehydrogenase protein which has strong formate de
CC	activity in the presence of an organic solvent. The mutant
CC	dehydrogenase protein of the invention may be used to pro
CC	using ketones as raw materials. Unlike prior formatted de
CC	proteins, the mutant polypeptides of the invention retain
CC	activity in the presence of the alcohol product and so pri
CC	product yield. The present sequence represents a Mycobact
CC	formate dehydrogenase protein mutant of the invention.
XX	
SQ	Sequence 401 AA;
	Query Match 1.1%; Score 8; DB 23; Length 401
	Best Local Similarity 100.0%; Pred. No. 42;
	Matches 8; Conservative 0; Mismatches 0; Indels
QY	172 VSGELGLR 179
Db	54 VSGELGLR 61
RESULT 32	
AAU99108	
ID	AAU99108 standard; Protein; 401 AA.
XX	
AC	AAU99108;
XX	
DT	24-SEP-2002 (first entry)
XX	
DE	Mycobacterium vaccae formate dehydrogenase C6A/C256S mutant
XX	
KW	Formate dehydrogenase; alcohol production; ketone;
KW	mutant; muten.
XX	
OS	Mycobacterium vaccae.
OS	Synthetic.
XX	

```

FH Key Location/Qualifiers
FT Misc-difference 6 /note= "Wild type Cys substituted by Ala"
FT Misc-difference 256 /note= "Wild type Cys substituted by Ser"
XX EP1211316-Al.
XX 05-JUN-2002.
XX 27-NOV-2001; 2001EP-0128170.
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX (DAIL ) DAICEL CHEM IND LTD.
XX Mitsuhashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX Claim 8; Page -; 42pp; English.
XX This invention relates to a mutant form of the Mycobacterium vaccae
XX formate dehydrogenase protein which has strong formate dehydrogenase
XX activity in the presence of an organic solvent. The mutant formate
XX dehydrogenase protein of the invention may be used to produce alcohols
XX using ketones as raw materials. Unlike prior formatted dehydrogenase,
XX proteins, the mutant polypeptides of the invention retain high enzymatic
XX activity in the presence of the alcohol product and so provides a higher
XX product yield. The present sequence represents a Mycobacterium vaccae
XX formate dehydrogenase protein mutant of the invention.
XX Sequence 401 AA;
XX
XX Query Match 1.1%; Score 8; DB 23; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 172 VSIGELGLR 179
XX |||||
XX Db 54 VSIGELGLR 61
XX
XX RESULT 33
XX AAU99109
XX ID AAU99109 standard; Protein; 401 AA.
XX AC AAU99109;
XX DT 24-SEP-2002 (first entry)
XX DE Mycobacterium vaccae formate dehydrogenase C5V/C256S mutant.
XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.
XX OS Mycobacterium vaccae.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 6 /note= "Wild type Cys substituted by Val"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ser"
XX EP1211316-Al.
XX 05-JUN-2002.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 6 /note= "Wild type Cys substituted by Val"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ser"
XX EP1211316-Al.
XX 05-JUN-2002.

```

```

XX 27-NOV-2001; 2001EP-0128170.
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX (DAIL ) DAICEL CHEM IND LTD.
XX Mitsuhashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX Claim 8; Page -; 42pp; English.
XX This invention relates to a mutant form of the Mycobacterium vaccae
XX formate dehydrogenase protein which has strong formate dehydrogenase
XX activity in the presence of an organic solvent. The mutant formate
XX dehydrogenase protein of the invention may be used to produce alcohols
XX using ketones as raw materials. Unlike prior formatted dehydrogenase,
XX proteins, the mutant polypeptides of the invention retain high enzymatic
XX activity in the presence of the alcohol product and so provides a higher
XX product yield. The present sequence represents a Mycobacterium vaccae
XX formate dehydrogenase protein mutant of the invention.
XX Sequence 401 AA;
XX
XX Query Match 1.1%; Score 8; DB 23; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 172 VSIGELGLR 179
XX |||||
XX Db 54 VSIGELGLR 61
XX
XX RESULT 34
XX AAU99110
XX ID AAU99110 standard; Protein; 401 AA.
XX AC AAU99110;
XX DT 24-SEP-2002 (first entry)
XX DE Mycobacterium vaccae formate dehydrogenase C6S/C256A mutant.
XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.
XX OS Mycobacterium vaccae.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 6 /note= "Wild type Cys substituted by Ser"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ala"
XX EP1211316-Al.
XX 05-JUN-2002.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX (DAIL ) DAICEL CHEM IND LTD.
XX Mitsuhashi K, Yamamoto H, Kimoto N;

```


Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
DB 54 VSGELGLR 61

RESULT 37

AAU99113
ID AAU99113 standard; Protein; 401 AA.

XX AC AAU99113;
XX

DT 24-SEP-2002 (first entry)

DE Mycobacterium vaccae formate dehydrogenase C256S mutant.

XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.

XX OS Mycobacterium vaccae.
XX Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 256

FT /note= "Wild type Cys substituted by Ser"

XX EP1211316-A1.

XX OS 05-JUN-2002.

XX PF 27-NOV-2001; 2001EP-0128170.

XX PR 29-NOV-2000; 2000JP-0363894.

XX PR 24-AUG-2001; 2001JP-0254631.

XX PA (DAIL) DAICEL CHEM IND LTD.

XX PI Mitsuhashi K, Yamamoto H, Kimoto N;

XX WPI; 2002-464925/50.

XX New mutant forms of formate dehydrogenase derived from Mycobacterium
PT vaccae have enhanced activity in the presence of organic solvents and
PT are useful for producing alcohols from ketone raw materials
XX
XX Claim 8; Page -; 42pp; English.

XX CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.

XX SQ Sequence 401 AA;

Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
DB 54 VSGELGLR 61

RESULT 38

AAU99114
ID AAU99114 standard; Protein; 401 AA.

XX AC AAU99114;

XX DT 24-SEP-2002 (first entry)

DE Mycobacterium vaccae formate dehydrogenase C146S/C256S mutant.

XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.

XX OS Mycobacterium vaccae.
XX Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 146

FT /note= "Wild type Cys substituted by Ser"

FT Misc-difference 256

FT /note= "Wild type Cys substituted by Ser"

XX EP1211316-A1.

XX PD 05-JUN-2002.

XX PF 27-NOV-2001; 2001EP-0128170.

XX PR 29-NOV-2000; 2000JP-0363894.

XX PR 24-AUG-2001; 2001JP-0254631.

XX PA (DAIL) DAICEL CHEM IND LTD.

XX PI Mitsuhashi K, Yamamoto H, Kimoto N;

XX WPI; 2002-464925/50.

XX New mutant forms of formate dehydrogenase derived from Mycobacterium
PT vaccae have enhanced activity in the presence of organic solvents and
PT are useful for producing alcohols from ketone raw materials
XX
XX Claim 8; Page -; 42pp; English.

XX CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.

XX SQ Sequence 401 AA;

Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
DB 54 VSGELGLR 61

RESULT 39

AAU99115
ID AAU99115 standard; Protein; 401 AA.

XX AC AAU99115;

XX DT 24-SEP-2002 (first entry)

DE Mycobacterium vaccae formate dehydrogenase C256V mutant.

```
KW Formate dehydrogenase; alcohol production; ketone;
KW mutant; mutin.
XX
XX Mycobacterium vaccae.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 256 /note= "Wild type Cys substituted by Val"
FT
XX
XX EP1211316-A1.
XX
XX 05-JUN-2002.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX
XX (DAIL ) DAICEL CHEM IND LTD.
XX
XX Mitsuhashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX
XX Claim 8; Page -; 42pp; English.
XX
XX This invention relates to a mutant form of the Mycobacterium vaccae
XX formate dehydrogenase protein which has strong formate dehydrogenase
XX activity in the presence of an organic solvent. The mutant formate
XX dehydrogenase protein of the invention may be used to produce alcohols
XX using ketones as raw materials. Unlike prior formatted dehydrogenase,
XX proteins, the mutant polypeptides of the invention retain high enzymatic
XX activity in the presence of the alcohol product and so provides a higher
XX product yield. The present sequence represents a Mycobacterium vaccae
XX formate dehydrogenase protein mutant of the invention.
XX
XX Sequence 401 AA;
XX
XX Query Match 1.1%; Score 8; DB 23; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 172 VSGELGLR 179
XX |||||
XX Db 54 VSGELGLR 61
XX
RESULT 40
AAU99116
ID AAU99116 standard; Protein; 401 AA.
XX
XX AC AAU99116;
XX
XX DT 24-SEP-2002 (first entry)
XX
XX DE Mycobacterium vaccae formate dehydrogenase C146S/C256V mutant.
XX
XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutin.
XX
XX OS Mycobacterium vaccae.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 146 /note= "Wild type Cys substituted by Ser"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Val"
XX FT
```

```
XX EP1211316-A1.
XX
XX PD 05-JUN-2002.
XX
XX PF 27-NOV-2001; 2001EP-0128170.
XX
XX PR 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX
XX (DAIL ) DAICEL CHEM IND LTD.
XX
XX PI Mitsuhashi K, Yamamoto H, Kimoto N;
XX DR WPI; 2002-464925/50.
XX
XX PT New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX
XX PS Claim 8; Page -; 42pp; English.
XX
XX CC This invention relates to a mutant form of the Mycobacterium vaccae
XX formate dehydrogenase protein which has strong formate dehydrogenase
XX activity in the presence of an organic solvent. The mutant formate
XX dehydrogenase protein of the invention may be used to produce alcohols
XX using ketones as raw materials. Unlike prior formatted dehydrogenase,
XX proteins, the mutant polypeptides of the invention retain high enzymatic
XX activity in the presence of the alcohol product and so provides a higher
XX product yield. The present sequence represents a Mycobacterium vaccae
XX formate dehydrogenase protein mutant of the invention.
XX
XX SQ Sequence 401 AA;
XX
XX Query Match 1.1%; Score 8; DB 23; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 172 VSGELGLR 179
XX |||||
XX Db 54 VSGELGLR 61
XX
Search completed: November 14, 2003, 11:06:54
Job time : 56 secs
```


101	6	0.8	31	4	US-09-106-075A-85	Sequence 85, Appl	174	6	0.8	147	2	US-08-647-960-7	Sequence 7, Appl
102	6	0.8	56	2	US-08-537-400-32	Sequence 32, Appl	175	6	0.8	148	6	5248606-46	Patent No. 5248606
103	6	0.8	59	4	US-08-469-260A-440	Sequence 440, App	176	6	0.8	149	4	US-09-252-991A-26802	Sequence 4, Appl
104	6	0.8	59	4	US-08-488-446-440	Sequence 440, App	177	6	0.8	150	4	US-09-373-953-4	Sequence 4, Appl
105	6	0.8	59	4	US-08-467-344A-440	Sequence 440, App	178	6	0.8	150	4	US-09-328-352-5004	Sequence 5004, App
106	6	0.8	62	1	US-08-210-266A-12	Sequence 12, Appl	179	6	0.8	150	4	US-08-911-844B-4	Sequence 4, Appl
107	6	0.8	62	1	US-08-688-673-12	Sequence 12, Appl	180	6	0.8	151	4	US-09-252-991A-23277	Sequence 23277, A
108	6	0.8	62	3	US-08-477-860C-12	Sequence 12, Appl	181	6	0.8	151	4	US-08-936-165A-264	Sequence 264, App
109	6	0.8	68	4	US-09-134-001C-5173	Sequence 5173, A	182	6	0.8	153	4	US-09-134-001C-3758	Sequence 3758, App
110	6	0.8	71	4	US-09-252-991A-28732	Sequence 28732, A	183	6	0.8	153	4	US-09-107-532A-6128	Sequence 6128, App
111	6	0.8	73	3	US-08-485-355B-46	Sequence 46, Appl	184	6	0.8	155	4	US-09-252-991A-20281	Sequence 20281, A
112	6	0.8	73	3	US-09-216-393B-293	Sequence 293, App	185	6	0.8	155	4	US-09-252-991A-32965	Sequence 32965, A
113	6	0.8	74	4	US-09-134-001C-5620	Sequence 5620, App	186	6	0.8	158	1	US-08-319-590-26	Sequence 26, Appl
114	6	0.8	80	4	US-09-252-991A-22470	Sequence 22470, A	187	6	0.8	158	1	US-08-487-001A-26	Sequence 26, Appl
115	6	0.8	81	2	US-08-245-511-14	Sequence 14, Appl	188	6	0.8	158	2	US-08-630-822A-26	Sequence 26, Appl
116	6	0.8	81	2	US-08-600-993A-14	Sequence 14, Appl	189	6	0.8	158	2	US-08-711-905-26	Sequence 26, Appl
117	6	0.8	84	4	US-09-252-991A-18487	Sequence 18487, A	190	6	0.8	158	2	US-09-005-069-26	Sequence 26, Appl
118	6	0.8	88	4	US-09-107-532A-6155	Sequence 6155, App	191	6	0.8	158	4	US-09-252-991A-32384	Sequence 32384, A
119	6	0.8	91	3	US-08-851-843A-220	Sequence 220, App	192	6	0.8	160	2	US-08-726-306A-35	Sequence 35, Appl
120	6	0.8	91	3	US-08-974-549A-339	Sequence 339, App	193	6	0.8	160	4	US-09-355-700-59	Sequence 59, Appl
121	6	0.8	91	3	US-08-854-050-220	Sequence 220, App	194	6	0.8	161	3	US-09-261-416-8	Sequence 8, Appl
122	6	0.8	91	4	US-09-430-323-220	Sequence 220, App	195	6	0.8	161	3	US-09-518-046-25	Sequence 25, Appl
123	6	0.8	92	3	US-09-074-658-78	Sequence 78, Appl	196	6	0.8	165	1	US-07-958-700B-106	Sequence 106, App
124	6	0.8	94	4	US-09-252-991A-23663	Sequence 23663, A	197	6	0.8	165	1	US-08-476-537-106	Sequence 106, App
125	6	0.8	97	4	US-09-134-001C-3197	Sequence 3197, App	198	6	0.8	165	1	US-08-485-607-106	Sequence 106, App
126	6	0.8	97	4	US-09-506-286B-55	Sequence 55, Appl	199	6	0.8	165	2	US-08-475-879-106	Sequence 106, App
127	6	0.8	98	1	US-09-053-945-4	Sequence 4, Appl	200	6	0.8	165	3	US-09-276-141-12	Sequence 12, Appl
128	6	0.8	100	6	5252466-19	Patent No. 5252466	201	6	0.8	165	3	US-09-276-141-13	Sequence 13, Appl
129	6	0.8	104	4	US-09-134-001C-4945	Sequence 4945, App	202	6	0.8	165	4	US-09-433-043B-106	Sequence 106, App
130	6	0.8	106	4	US-09-252-991A-20137	Sequence 20137, A	203	6	0.8	167	4	US-09-252-991A-26054	Sequence 26054, App
131	6	0.8	106	4	US-09-328-352-5211	Sequence 5211, App	204	6	0.8	167	4	US-09-252-991A-29214	Sequence 29214, A
132	6	0.8	108	4	US-09-134-001C-3747	Sequence 3747, App	205	6	0.8	169	1	US-08-534-975-2	Sequence 2, Appl
133	6	0.8	109	4	US-09-227-357-462	Sequence 462, App	206	6	0.8	169	2	US-08-954-470-2	Sequence 2, Appl
134	6	0.8	112	4	US-09-341-461-34	Sequence 34, Appl	207	6	0.8	169	3	US-09-129-855A-2	Sequence 2, Appl
135	6	0.8	115	4	US-09-376-330-22	Sequence 22, Appl	208	6	0.8	169	3	US-09-247-154-2	Sequence 2, Appl
136	6	0.8	116	4	US-09-732-210-1691	Sequence 1691, App	209	6	0.8	169	4	US-09-480-783-2	Sequence 2, Appl
137	6	0.8	117	4	US-09-732-210-772	Sequence 772, App	210	6	0.8	169	4	US-09-610-813-2	Sequence 2, Appl
138	6	0.8	119	4	US-09-107-532A-7227	Sequence 7227, App	211	6	0.8	169	4	US-09-129-855A-2	Sequence 2, Appl
139	6	0.8	120	4	US-09-660-552-31	Sequence 31, Appl	212	6	0.8	170	4	US-09-433-043B-127	Sequence 127, App
140	6	0.8	120	4	US-08-311-731A-338	Sequence 338, App	213	6	0.8	170	4	US-09-252-991A-20681	Sequence 20681, A
141	6	0.8	121	4	US-09-199-637A-421	Sequence 421, App	214	6	0.8	170	4	US-09-551-921-14	Sequence 14, Appl
142	6	0.8	121	4	US-09-328-352-5325	Sequence 5325, App	215	6	0.8	174	3	US-08-911-853-19	Sequence 19, Appl
143	6	0.8	122	4	US-08-392-459-30	Sequence 30, Appl	216	6	0.8	174	3	US-09-479-409-19	Sequence 19, Appl
144	6	0.8	122	5	PCT-US91-08525-30	Sequence 30, Appl	217	6	0.8	174	3	US-09-153-586-23	Sequence 23, Appl
145	6	0.8	123	1	PCT-US93-04384-10	Sequence 10, Appl	218	6	0.8	174	4	US-09-479-453-19	Sequence 19, Appl
146	6	0.8	123	1	US-09-053-945-2	Sequence 2, Appl	219	6	0.8	175	1	US-07-783-705A-6	Sequence 6, Appl
147	6	0.8	123	1	US-08-061-314A-2	Sequence 2, Appl	220	6	0.8	175	1	US-08-624-135-7	Sequence 7, Appl
148	6	0.8	123	4	US-09-328-352-4586	Sequence 4586, App	221	6	0.8	175	3	US-08-129-030-22	Sequence 22, Appl
149	6	0.8	125	3	PCT-US91-08525-54	Sequence 54, Appl	222	6	0.8	175	4	US-09-937-155-7	Sequence 7, Appl
150	6	0.8	125	5	PCT-US91-08525-54	Sequence 54, Appl	223	6	0.8	176	4	US-09-252-991A-28483	Sequence 28483, A
151	6	0.8	126	5	PCT-US93-04384-14	Sequence 14, Appl	224	6	0.8	177	4	US-09-199-637A-377	Sequence 377, App
152	6	0.8	126	4	US-09-732-210-269	Sequence 269, App	225	6	0.8	177	4	US-09-252-991A-25925	Sequence 25925, A
153	6	0.8	129	4	US-09-107-532A-5086	Sequence 5086, App	226	6	0.8	182	4	US-09-252-991A-28591	Sequence 28591, A
154	6	0.8	130	4	US-09-485-316A-11	Sequence 11, Appl	227	6	0.8	182	4	US-08-328-352-7019	Sequence 7019, App
155	6	0.8	130	4	US-09-485-316A-12	Sequence 12, Appl	228	6	0.8	183	4	US-09-373-953-2	Sequence 2, Appl
156	6	0.8	130	4	US-09-485-316A-13	Sequence 13, Appl	229	6	0.8	183	4	US-08-911-844B-2	Sequence 2, Appl
157	6	0.8	133	4	US-09-461-325-409	Sequence 409, App	230	6	0.8	185	4	US-09-134-001C-4741	Sequence 4741, App
158	6	0.8	133	4	US-09-252-991A-27120	Sequence 27120, App	231	6	0.8	185	4	US-09-134-001C-4741	Sequence 4741, App
159	6	0.8	136	3	US-09-276-141-8	Sequence 8, Appl	232	6	0.8	188	4	US-09-393-627B-19	Sequence 19, Appl
160	6	0.8	136	3	US-09-276-141-14	Sequence 14, Appl	233	6	0.8	188	4	US-09-489-847-314	Sequence 314, App
161	6	0.8	137	1	US-08-258-026A-18	Sequence 18, Appl	234	6	0.8	190	3	US-08-392-459-20	Sequence 20, Appl
162	6	0.8	137	4	US-09-252-991A-22927	Sequence 22927, A	235	6	0.8	190	5	PCT-US91-08525-20	Sequence 20, Appl
163	6	0.8	137	5	PCT-US95-07541-18	Sequence 18, Appl	236	6	0.8	190	4	US-08-936-165A-465	Sequence 465, App
164	6	0.8	141	4	US-09-252-991A-19268	Sequence 19268, A	237	6	0.8	194	4	US-09-252-991A-23770	Sequence 23770, A
165	6	0.8	141	4	US-09-252-991A-19815	Sequence 19815, A	238	6	0.8	195	4	US-05-585-173B-18	Sequence 18, Appl
166	6	0.8	141	4	US-09-252-991A-25082	Sequence 25082, A	239	6	0.8	197	2	US-08-356-361-27	Sequence 27, Appl
167	6	0.8	141	4	US-09-252-991A-25325	Sequence 25325, A	240	6	0.8	199	4	US-08-769-967A-27	Sequence 27, Appl
168	6	0.8	142	4	US-09-252-991A-19283	Sequence 19283, A	241	6	0.8	199	4	US-09-252-991A-20230	Sequence 20230, A
169	6	0.8	143	4	US-09-252-991A-17861	Sequence 17861, A	242	6	0.8	199	4	US-09-252-991A-26642	Sequence 26642, A
170	6	0.8	143	4	US-09-252-991A-21044	Sequence 21044, A	243	6	0.8	201	4	US-09-198-452A-595	Sequence 595, App
171	6	0.8	144	4	US-09-252-991A-21233	Sequence 21233, A	244	6	0.8	203	4	US-09-328-352-4960	Sequence 4960, App
172	6	0.8	144	4	US-09-252-991A-26540	Sequence 26540, A	245	6	0.8	204	4	US-09-252-991A-19193	Sequence 19193, A
173	6	0.8	146	4	US-09-660-552-29	Sequence 29, Appl	246	6	0.8	206	4	US-09-252-991A-29093	Sequence 29093, A

247 6 0.8 209 4 US-09-252-991A-19943 Sequence 19943, A
 248 6 0.8 209 4 US-09-252-991A-25767 Sequence 25767, A
 249 6 0.8 213 4 US-09-252-991A-22098 Sequence 22098, A
 250 6 0.8 215 4 US-09-252-991A-5383A-5 Sequence 5, Appli
 251 6 0.8 216 4 US-09-229-583A-6 Sequence 3, Appli
 252 6 0.8 217 4 US-09-229-583A-6 Sequence 6, Appli
 253 6 0.8 219 3 US-08-834-776A-3 Sequence 3, Appli
 254 6 0.8 219 4 US-09-328-352-7187 Sequence 7187, Ap
 255 6 0.8 220 4 US-09-107-532A-6317 Sequence 6317, Ap
 256 6 0.8 225 4 US-09-252-991A-32865 Sequence 32865, A
 257 6 0.8 225 4 US-09-634-238A-289 Sequence 289, App
 258 6 0.8 227 4 US-09-252-991A-19667 Sequence 19667, A
 259 6 0.8 227 4 US-09-252-991A-21816 Sequence 21816, A
 260 6 0.8 227 4 US-09-252-991A-31689 Sequence 31689, A
 261 6 0.8 228 4 US-09-431-888-7 Sequence 7, Appli
 262 6 0.8 229 4 US-09-252-991A-21868 Sequence 21868, A
 263 6 0.8 230 4 US-09-506-286B-51 Sequence 51, Appl
 264 6 0.8 230 4 US-09-506-286B-58 Sequence 58, Appl
 265 6 0.8 230 4 US-09-506-286B-58 Sequence 58, Appl
 266 6 0.8 231 4 US-09-252-991A-17396 Sequence 17396, A
 267 6 0.8 231 4 US-09-530-423-2 Sequence 2, Appli
 268 6 0.8 232 4 US-09-328-352-4900 Sequence 4900, Ap
 269 6 0.8 233 3 US-09-248-335-60 Sequence 60, Appl
 270 6 0.8 233 4 US-09-252-991A-17686 Sequence 17686, A
 271 6 0.8 236 3 US-09-248-335-68 Sequence 68, Appl
 272 6 0.8 238 4 US-09-605-858B-35 Sequence 35, Appl
 273 6 0.8 238 4 US-09-071-035-156 Sequence 156, App
 274 6 0.8 242 4 US-09-252-991A-19136 Sequence 19136, A
 275 6 0.8 244 2 US-08-463-911-7 Sequence 7, Appli
 276 6 0.8 244 3 US-09-140-804-3 Sequence 3, Appli
 277 6 0.8 244 4 US-09-336-536-20 Sequence 20, Appl
 278 6 0.8 244 4 US-09-530-423-1 Sequence 1, Appli
 279 6 0.8 244 4 US-09-686-838B-3 Sequence 3, Appli
 280 6 0.8 244 4 US-09-911-176B-48 Sequence 48, Appl
 281 6 0.8 244 4 US-09-552-225A-3 Sequence 3, Appli
 282 6 0.8 244 4 US-09-619-740-51 Sequence 51, Appl
 283 6 0.8 244 4 US-09-252-991A-29399 Sequence 29399, A
 284 6 0.8 244 4 US-09-776-976-6 Sequence 6, Appli
 285 6 0.8 244 4 US-09-909-547-6 Sequence 6, Appli
 286 6 0.8 244 4 US-09-569-852B-6 Sequence 6, Appli
 287 6 0.8 246 4 US-09-252-991A-20050 Sequence 20050, A
 288 6 0.8 247 2 US-08-463-911-2 Sequence 2, Appli
 289 6 0.8 247 3 US-09-140-804-8 Sequence 8, Appli
 290 6 0.8 247 3 US-09-118-408-3 Sequence 3, Appli
 291 6 0.8 247 4 US-09-506-855-3 Sequence 3, Appli
 292 6 0.8 247 4 US-09-686-838B-8 Sequence 8, Appli
 293 6 0.8 247 4 US-09-911-176B-3 Sequence 3, Appli
 294 6 0.8 247 4 US-09-619-740-3 Sequence 3, Appli
 295 6 0.8 247 4 US-09-252-991A-32109 Sequence 32109, A
 296 6 0.8 247 4 US-09-776-976-2 Sequence 2, Appli
 297 6 0.8 247 4 US-09-776-976-4 Sequence 4, Appli
 298 6 0.8 247 4 US-09-506-852-3 Sequence 3, Appli
 299 6 0.8 247 4 US-09-909-547-2 Sequence 2, Appli
 300 6 0.8 247 4 US-09-909-547-4 Sequence 4, Appli
 301 6 0.8 248 4 US-09-252-991A-22310 Sequence 22310, A
 302 6 0.8 250 4 US-09-331-568A-6 Sequence 6, Appli
 303 6 0.8 251 3 US-09-175-014-2 Sequence 2, Appli
 304 6 0.8 252 1 US-08-271-354-14 Sequence 14, Appl
 305 6 0.8 252 2 US-08-565-861-14 Sequence 14, Appl
 306 6 0.8 252 5 PCT-US94-0765B-14 Sequence 14, Appl
 307 6 0.8 254 1 US-08-378-761A-11 Sequence 11, Appl
 308 6 0.8 254 1 US-08-485-286-11 Sequence 11, Appl
 309 6 0.8 254 2 US-08-356-361-29 Sequence 29, Appl
 310 6 0.8 254 2 US-08-356-361-30 Sequence 30, Appl
 311 6 0.8 254 2 US-08-769-967A-29 Sequence 29, Appl
 312 6 0.8 254 2 US-08-769-967A-30 Sequence 30, Appl
 313 6 0.8 254 3 US-09-013-881-8 Sequence 8, Appli
 314 6 0.8 254 4 US-09-612-473-8 Sequence 8, Appli
 315 6 0.8 255 3 US-08-944-483-67 Sequence 67, Appl
 316 6 0.8 255 4 US-09-252-991A-17685 Sequence 17685, A
 317 6 0.8 255 4 US-09-252-991A-28094 Sequence 28094, A
 318 6 0.8 256 2 US-09-027-337-3 Sequence 3, Appli
 319 6 0.8 256 2 US-08-484-905-114 Sequence 114, App

320 6 0.8 256 3 US-08-481-985B-114 Sequence 114, App
 321 6 0.8 256 3 US-08-370-476-114 Sequence 114, App
 322 6 0.8 256 4 US-09-644-600-3 Sequence 3, Appli
 323 6 0.8 256 4 US-09-252-991A-19714 Sequence 19714, A
 324 6 0.8 256 4 US-09-252-991A-21011 Sequence 21011, A
 325 6 0.8 257 1 US-08-378-761A-13 Sequence 13, Appl
 326 6 0.8 257 1 US-08-485-286-13 Sequence 13, Appl
 327 6 0.8 258 3 US-08-737-226-6 Sequence 6, Appli
 328 6 0.8 258 3 US-09-097-767A-6 Sequence 6, Appli
 329 6 0.8 261 1 US-08-378-761A-9 Sequence 9, Appli
 330 6 0.8 261 1 US-08-485-286-9 Sequence 9, Appli
 331 6 0.8 262 6 5248606-43 Patent No. 5248606
 332 6 0.8 263 4 US-09-252-991A-28494 Sequence 28494, A
 333 6 0.8 263 1 US-08-378-761A-21 Sequence 21, Appl
 334 6 0.8 265 1 US-08-485-286-21 Sequence 21, Appl
 335 6 0.8 265 2 US-08-663-566A-50 Sequence 50, Appl
 336 6 0.8 266 2 US-08-288-065A-50 Sequence 50, Appl
 337 6 0.8 266 2 US-08-362-240A-50 Sequence 50, Appl
 338 6 0.8 266 2 US-08-804-372A-14 Sequence 14, Appl
 339 6 0.8 266 4 US-09-134-001C-5453 Sequence 5453, Ap
 340 6 0.8 266 5 PCT-US95-10245-50 Sequence 50, Appl
 341 6 0.8 266 4 US-09-071-035-154 Sequence 154, App
 342 6 0.8 267 4 US-09-252-991A-27950 Sequence 27950, A
 343 6 0.8 268 4 US-08-362-525-8 Sequence 8, Appli
 344 6 0.8 270 3 US-09-252-991A-22707 Sequence 22707, A
 345 6 0.8 272 4 US-08-137-175A-5 Sequence 5, Appli
 346 6 0.8 273 1 US-08-479-017-5 Sequence 5, Appli
 347 6 0.8 273 3 US-09-196-293-13 Sequence 13, Appl
 348 6 0.8 273 3 US-08-209-603B-13 Sequence 13, Appl
 349 6 0.8 273 4 US-09-711-546-13 Sequence 13, Appl
 350 6 0.8 273 4 US-09-684-385-10 Sequence 10, Appl
 351 6 0.8 273 6 5248606-41 Patent No. 5248606
 352 6 0.8 275 1 US-08-378-761A-7 Sequence 7, Appli
 353 6 0.8 276 1 US-08-485-286-7 Sequence 7, Appli
 354 6 0.8 276 2 US-08-852-743-7 Sequence 7, Appli
 355 6 0.8 276 3 US-09-185-370-7 Sequence 7, Appli
 356 6 0.8 276 3 US-07-941-651-2 Sequence 2, Appli
 357 6 0.8 278 1 US-08-279-996-2 Sequence 2, Appli
 358 6 0.8 278 4 US-09-252-991A-21844 Sequence 21844, A
 359 6 0.8 278 3 US-09-097-767A-10 Sequence 10, Appl
 360 6 0.8 279 4 US-09-252-991A-32422 Sequence 32422, A
 361 6 0.8 280 4 US-09-252-991A-31168 Sequence 31168, A
 362 6 0.8 285 4 US-09-252-991A-29826 Sequence 29826, A
 363 6 0.8 285 6 5248606-45 Patent No. 5248606
 364 6 0.8 286 1 US-08-378-761A-5 Sequence 5, Appli
 365 6 0.8 286 1 US-08-485-286-5 Sequence 5, Appli
 366 6 0.8 286 4 US-09-252-991A-32274 Sequence 32274, A
 367 6 0.8 286 4 US-09-252-991A-23449 Sequence 23449, A
 368 6 0.8 288 4 US-09-007-288B-145 Sequence 145, App
 369 6 0.8 289 4 US-08-311-731A-396 Sequence 396, Appl
 370 6 0.8 289 4 US-09-358-383C-8 Sequence 8, Appli
 371 6 0.8 290 4 US-09-252-991A-18740 Sequence 18740, A
 372 6 0.8 290 1 US-08-434-255-13 Sequence 13, Appl
 373 6 0.8 291 1 US-08-459-967-13 Sequence 13, Appl
 374 6 0.8 291 1 US-08-460-327-13 Sequence 13, Appl
 375 6 0.8 291 1 US-08-459-871-13 Sequence 13, Appl
 376 6 0.8 291 1 US-08-244-686-4 Sequence 4, Appli
 377 6 0.8 291 2 US-08-784-651-16 Sequence 16, Appl
 378 6 0.8 291 2 US-08-321-426-10 Sequence 10, Appl
 379 6 0.8 291 2 US-08-479-275D-2 Sequence 2, Appli
 380 6 0.8 291 2 US-08-488-271B-2 Sequence 2, Appli
 381 6 0.8 291 2 US-08-140-008A-2 Sequence 2, Appli
 382 6 0.8 291 2 US-08-701-339-2 Sequence 2, Appli
 383 6 0.8 291 3 US-08-816-915-10 Sequence 10, Appl
 384 6 0.8 291 3 US-09-024-532-6 Sequence 6, Appli
 385 6 0.8 291 4 US-09-007-288E-16 Sequence 16, Appl
 386 6 0.8 291 5 PCT-US95-07743-10 Sequence 10, Appl
 387 6 0.8 292 4 US-09-252-991A-22264 Sequence 22264, A
 388 6 0.8 292 4 US-09-252-991A-19896 Sequence 19896, A
 389 6 0.8 293 4 US-09-007-288B-142 Sequence 142, App
 390 6 0.8 294 4 US-09-007-288B-143 Sequence 143, App
 391 6 0.8 294 4 US-09-252-991A-33135 Sequence 33135, A

393	6	0.8	295	4	US-09-724-623-115	Sequence 115, Appl	466	367	4	US-09-252-991A-17216	Sequence 17216, A
394	6	0.8	297	4	US-09-252-991A-22847	Sequence 22847, A	467	368	4	US-09-252-991A-20724	Sequence 20724, A
395	6	0.8	298	4	US-09-134-001C-3317	Sequence 3317, A	468	369	4	US-09-816-088-4	Sequence 4, Appl
396	6	0.8	298	4	US-09-252-991A-29045	Sequence 29045, A	469	369	4	US-09-634-955B-8	Sequence 8, Appl
397	6	0.8	299	4	US-09-107-532A-5109	Sequence 5109, Appl	470	370	1	US-07-662-007B-37	Sequence 37, Appl
398	6	0.8	301	1	US-08-378-761A-2	Sequence 2, Appl	471	370	1	US-07-824-247-37	Sequence 37, Appl
399	6	0.8	301	1	US-08-485-286-2	Sequence 2, Appl	472	370	3	US-08-470-204A-37	Sequence 37, Appl
400	6	0.8	301	6	5248606-2	Patent No. 5248606	473	371	4	US-09-252-991A-17066	Sequence 17066, A
401	6	0.8	302	4	US-09-252-991A-24052	Sequence 24052, A	474	372	4	US-09-252-991A-18377	Sequence 18377, A
402	6	0.8	302	4	US-09-252-991A-24382	Sequence 24382, A	475	374	4	US-09-679-279-6	Sequence 6, Appl
403	6	0.8	303	4	US-09-328-352-7045	Sequence 7045, Appl	476	374	4	US-09-107-532A-5084	Sequence 5084, Appl
404	6	0.8	308	4	US-09-328-352-7158	Sequence 7158, Appl	477	375	4	US-09-252-991A-27411	Sequence 27411, A
405	6	0.8	308	4	US-09-107-532A-6023	Sequence 6023, Appl	478	376	4	US-09-820-002-2	Sequence 2, Appl
406	6	0.8	311	4	US-09-252-991A-27736	Sequence 27736, A	479	378	4	US-09-252-991A-23815	Sequence 23815, A
407	6	0.8	312	4	US-09-107-532A-4496	Sequence 4496, Appl	480	379	3	US-09-191-136-32	Sequence 32, Appl
408	6	0.8	313	4	US-09-252-991A-22500	Sequence 22500, A	481	380	2	US-08-227-108-16	Sequence 16, Appl
409	6	0.8	314	4	US-09-328-352-4934	Sequence 4934, Appl	482	380	2	US-08-073-674-16	Sequence 16, Appl
410	6	0.8	315	1	US-08-378-761A-15	Sequence 15, Appl	483	381	4	US-09-107-532A-4570	Sequence 4570, Appl
411	6	0.8	315	1	US-08-485-286-15	Sequence 15, Appl	484	381	4	US-09-216-393B-8	Sequence 8, Appl
412	6	0.8	315	4	US-09-252-991A-17051	Sequence 17051, A	485	382	1	US-07-824-247-42	Sequence 42, Appl
413	6	0.8	316	4	US-09-252-991A-19073	Sequence 19073, A	486	382	1	US-08-142-473A-2	Sequence 2, Appl
414	6	0.8	316	4	US-09-252-991A-29585	Sequence 29585, A	487	382	1	US-08-464-523B-26	Sequence 26, Appl
415	6	0.8	319	4	US-09-393-627B-4	Sequence 4, Appl	488	382	1	US-08-469-203A-2	Sequence 2, Appl
416	6	0.8	320	4	US-09-252-991A-23385	Sequence 23385, A	489	382	1	US-08-469-203A-2	Sequence 2, Appl
417	6	0.8	320	4	US-09-252-991A-31763	Sequence 31763, A	490	382	2	US-08-605-106-12	Sequence 12, Appl
418	6	0.8	321	4	US-08-909-125-5	Sequence 5, Appl	491	382	2	US-08-440-845D-1	Sequence 1, Appl
419	6	0.8	321	4	US-09-328-352-7164	Sequence 7164, Appl	492	382	3	US-08-470-204A-42	Sequence 42, Appl
420	6	0.8	322	4	US-09-559-023-2	Sequence 2, Appl	493	382	3	US-08-868-458-1	Sequence 1, Appl
421	6	0.8	323	4	US-09-252-991A-23289	Sequence 23289, A	494	382	4	US-09-252-991A-29879	Sequence 29879, A
422	6	0.8	324	3	US-08-946-914-11	Sequence 11, Appl	495	383	4	US-09-634-238-277	Sequence 277, Appl
423	6	0.8	324	4	US-09-370-838-45	Sequence 45, Appl	496	383	4	US-09-252-991A-18162	Sequence 18162, A
424	6	0.8	324	4	US-09-656-450-11	Sequence 11, Appl	497	383	4	US-09-252-991A-27087	Sequence 27087, A
425	6	0.8	325	4	US-09-252-991A-18378	Sequence 18378, A	498	384	4	US-09-252-991A-28606	Sequence 28606, A
426	6	0.8	326	4	US-09-252-991A-22024	Sequence 22024, A	499	384	4	US-09-252-991A-28606	Sequence 28606, A
427	6	0.8	326	4	US-09-252-991A-30438	Sequence 30438, A	500	385	4	US-09-107-532A-5168	Sequence 5168, A
428	6	0.8	328	4	US-09-252-991A-28130	Sequence 28130, A	501	385	4	US-09-252-991A-22751	Sequence 22751, A
429	6	0.8	330	3	US-09-267-031-12	Sequence 12, Appl	502	388	4	US-09-107-532A-5369	Sequence 5369, A
430	6	0.8	330	3	US-09-267-031-16	Sequence 16, Appl	503	388	4	US-09-252-991A-24858	Sequence 24858, A
431	6	0.8	332	4	US-09-134-001C-4032	Sequence 4032, Appl	504	390	4	US-09-127-124-1	Sequence 1, Appl
432	6	0.8	332	4	US-09-252-991A-29873	Sequence 29873, A	505	393	4	US-08-979-170-2	Sequence 2, Appl
433	6	0.8	333	3	US-09-267-031-4	Sequence 4, Appl	506	393	4	US-09-842-898-2	Sequence 2, Appl
434	6	0.8	333	4	US-09-252-991A-25991	Sequence 25991, A	507	395	4	US-09-252-991A-24596	Sequence 24596, A
435	6	0.8	334	3	US-09-267-031-2	Sequence 2, Appl	508	395	4	US-09-107-532A-6955	Sequence 6955, A
436	6	0.8	337	4	US-09-328-352-7830	Sequence 7830, Appl	509	396	3	US-08-818-112-111	Sequence 111, Appl
437	6	0.8	338	2	US-08-441-857-10	Sequence 10, Appl	510	396	4	US-08-818-111-106	Sequence 106, Appl
438	6	0.8	338	3	US-08-193-159-10	Sequence 10, Appl	511	396	4	US-09-056-556-111	Sequence 111, Appl
439	6	0.8	338	4	US-09-330-611-6	Sequence 6, Appl	512	396	4	US-09-072-556-106	Sequence 106, Appl
440	6	0.8	338	4	US-09-252-991A-22435	Sequence 22435, A	513	396	4	US-09-252-991A-28379	Sequence 28379, A
441	6	0.8	339	4	US-09-252-991A-28546	Sequence 28546, A	514	396	4	US-09-252-991A-32164	Sequence 32164, A
442	6	0.8	339	4	US-09-252-991A-26507	Sequence 26507, A	515	397	4	US-09-252-991A-20668	Sequence 20668, A
443	6	0.8	339	4	US-09-107-532A-4033	Sequence 4033, Appl	516	397	4	US-09-252-991A-31874	Sequence 31874, A
444	6	0.8	340	4	US-09-252-991A-24822	Sequence 24822, A	517	398	4	US-09-348-930A-9	Sequence 9, Appl
445	6	0.8	342	3	US-09-330-611-4	Sequence 4, Appl	518	402	1	US-08-378-761A-17	Sequence 17, Appl
446	6	0.8	346	4	US-09-328-352-5541	Sequence 5541, Appl	519	402	1	US-08-485-286-17	Sequence 17, Appl
447	6	0.8	348	3	US-09-113-536-2	Sequence 2, Appl	520	402	3	US-08-861-774E-14	Sequence 14, Appl
448	6	0.8	348	4	US-09-624-183-2	Sequence 2, Appl	521	403	4	US-09-503-391-8	Sequence 8, Appl
449	6	0.8	348	4	US-09-252-991A-20593	Sequence 20593, A	522	403	4	US-09-252-991A-29379	Sequence 29379, A
450	6	0.8	348	4	US-09-252-991A-23148	Sequence 23148, A	523	404	4	US-09-252-991A-23011	Sequence 23011, A
451	6	0.8	348	5	PCT-US85-05785-2	Sequence 23148, A	524	404	4	US-09-107-532A-6414	Sequence 6414, A
452	6	0.8	350	2	US-08-999-811-4	Sequence 2, Appl	525	405	4	US-09-252-991A-23678	Sequence 23678, A
453	6	0.8	350	2	US-08-824-996-2	Sequence 4, Appl	526	406	4	US-09-252-991A-21303	Sequence 21303, A
454	6	0.8	350	3	US-09-042-105-4	Sequence 2, Appl	527	408	4	US-09-252-991A-23711	Sequence 23711, A
455	6	0.8	350	3	US-08-510-133A-33	Sequence 4, Appl	528	411	4	US-09-252-991A-27927	Sequence 27927, A
456	6	0.8	350	3	US-08-585-895-33	Sequence 33, Appl	529	411	4	US-09-311-731A-369	Sequence 369, Appl
457	6	0.8	353	4	US-09-252-991A-22306	Sequence 22306, A	530	412	4	US-09-328-352-8136	Sequence 8136, A
458	6	0.8	355	4	US-09-252-991A-18900	Sequence 18900, A	531	415	4	US-09-252-991A-31846	Sequence 31846, A
459	6	0.8	358	4	US-09-328-352-4459	Sequence 4459, Appl	532	417	4	US-09-820-002-4	Sequence 4, Appl
460	6	0.8	360	4	US-09-364-230-4	Sequence 4, Appl	533	417	4	US-09-252-991A-16876	Sequence 16876, A
461	6	0.8	360	4	US-09-107-532A-6666	Sequence 6666, Appl	534	419	2	US-08-999-811-2	Sequence 2, Appl
462	6	0.8	362	4	US-09-252-991A-20079	Sequence 20079, A	535	419	3	US-09-042-105-2	Sequence 2, Appl
463	6	0.8	362	4	US-09-512-346-2	Sequence 2, Appl	536	419	3	US-09-042-105-18	Sequence 18, Appl
464	6	0.8	363	4	US-09-252-991A-26741	Sequence 26741, A	537	419	3	US-08-795-430-8	Sequence 8, Appl
465	6	0.8	365	4	US-09-252-991A-30595	Sequence 30595, A	538	419	3	US-08-510-133A-35	Sequence 35, Appl

539	6	0.8	4	US-09-355-700-8	Sequence 8, Appli	612	4	US-09-252-991A-24392	Sequence 24392, A
540	6	0.8	4	US-09-355-700-58	Sequence 58, Appli	613	4	US-09-252-991A-17093	Sequence 17093, A
541	6	0.8	4	US-09-134-001C-3441	Sequence 3441, Ap	614	4	US-09-252-991A-131455	Sequence 131455, A
542	6	0.8	4	US-08-601-132-33	Sequence 33, Appli	615	6	US-09-198-452A-81	Sequence 81, Appl
543	6	0.8	4	US-08-706-054A-3	Sequence 3, Appli	616	6	US-09-724-623-77	Sequence 77, Appl
544	6	0.8	4	US-09-328-352-6670	Sequence 6670, Ap	617	6	US-08-976-063E-28	Sequence 2, Appli
545	6	0.8	4	US-09-313-293-3	Sequence 3, Appli	618	6	US-09-557-921-2	Sequence 2, Appli
546	6	0.8	4	PCT-US96-09001-2	Sequence 2, Appli	619	6	US-09-252-991A-25334	Sequence 19, Appl
547	6	0.8	4	US-09-134-001C-5356	Sequence 5356, Ap	620	6	US-08-378-761A-19	Sequence 19, Appl
548	6	0.8	4	US-09-107-532A-6341	Sequence 6341, Ap	621	6	US-08-485-286-19	Sequence 4, Appli
549	6	0.8	4	US-08-700-749A-5	Sequence 5, Appli	622	6	US-08-553-436A-4	Sequence 19230, A
550	6	0.8	4	US-09-020-684-5	Sequence 5, Appli	623	6	US-08-252-991A-19230	Sequence 23, Appl
551	6	0.8	4	US-09-020-467-5	Sequence 5, Appli	624	6	US-08-378-761A-23	Sequence 23, Appl
552	6	0.8	4	US-09-020-685-5	Sequence 5, Appli	625	6	US-08-485-286-23	Patent No. 5486473
553	6	0.8	4	US-09-020-683-5	Sequence 5, Appli	626	6	US-08-485-286-23	Sequence 28162, A
554	6	0.8	4	US-09-252-991A-23349	Sequence 23349, A	627	6	US-09-252-991A-28162	Sequence 28162, A
555	6	0.8	4	US-08-403-797-2	Sequence 2, Appli	628	6	US-09-252-991A-29568	Sequence 29568, A
556	6	0.8	4	US-09-328-352-4392	Sequence 4392, Ap	629	6	US-09-107-532A-4616	Sequence 4616, Ap
557	6	0.8	4	US-09-252-991A-31909	Sequence 31909, A	630	6	US-09-252-991A-18725	Sequence 18725, A
558	6	0.8	4	US-09-188-930-341	Sequence 341, App	631	6	US-09-172-952-26	Sequence 26, Appl
559	6	0.8	4	US-09-381-681-3	Sequence 3, Appli	632	6	US-09-252-991A-28182	Sequence 28182, A
560	6	0.8	4	US-09-685-462-10	Sequence 10, Appl	633	6	US-09-252-991A-23848	Sequence 23848, A
561	6	0.8	4	US-09-312-283C-341	Sequence 341, App	634	6	US-09-325-932A-149	Sequence 149, App
562	6	0.8	4	US-09-252-991A-22739	Sequence 22739, A	635	6	US-09-172-952-15	Sequence 15, Appl
563	6	0.8	4	US-09-134-001C-3665	Sequence 3665, Ap	636	6	US-08-378-761A-25	Sequence 25, Appl
564	6	0.8	4	US-09-252-991A-31231	Sequence 31231, A	637	6	US-08-485-286-25	Sequence 25, Appl
565	6	0.8	4	US-09-252-991A-23489	Sequence 23489, A	638	6	US-09-499-302A-7	Sequence 7, Appli
566	6	0.8	4	US-09-252-991A-20874	Sequence 20874, A	639	6	US-08-700-651-6	Sequence 6, Appli
567	6	0.8	4	US-09-252-991A-17213	Sequence 17213, A	640	6	US-09-588-995A-66	Sequence 66, Appl
568	6	0.8	4	US-09-252-991A-17836	Sequence 17836, A	641	6	US-08-123-161A-14	Sequence 14, Appl
569	6	0.8	4	US-09-252-991A-17836	Sequence 17836, A	642	6	US-08-483-278-14	Sequence 14, Appl
570	6	0.8	4	US-08-353-638-9	Sequence 9, Appli	643	6	US-09-328-352-5823	Sequence 5823, Ap
571	6	0.8	4	US-09-160-036-1	Sequence 1, Appli	644	6	US-09-134-001C-4383	Sequence 4383, Ap
572	6	0.8	4	US-09-252-991A-21718	Sequence 21718, A	645	6	US-09-252-991A-30561	Sequence 30561, A
573	6	0.8	4	US-09-328-352-7336	Sequence 7336, Ap	646	6	US-08-097-829-6	Sequence 6, Appli
574	6	0.8	4	US-09-191-136-31	Sequence 31, Appl	647	6	US-08-577-403-6	Sequence 6, Appli
575	6	0.8	4	US-09-252-991A-18870	Sequence 18870, A	648	6	US-08-980-523-9	Sequence 9, Appli
576	6	0.8	4	US-09-252-991A-21563	Sequence 21563, A	649	6	US-09-252-991A-26154	Sequence 26154, A
577	6	0.8	4	US-09-188-930-296	Sequence 296, App	650	6	US-09-252-991A-27294	Sequence 27294, A
578	6	0.8	4	US-09-312-283C-296	Sequence 296, App	651	6	US-09-102-528-21	Sequence 21, Appl
579	6	0.8	4	US-09-312-283C-406	Sequence 406, App	652	6	US-09-252-991A-26146	Sequence 26146, A
580	6	0.8	4	US-09-310-463-18	Sequence 18, Appl	653	6	US-08-924-183-8	Sequence 8, Appli
581	6	0.8	4	US-08-842-248A-18	Sequence 18, Appl	654	6	US-09-488-364-8	Sequence 8, Appli
582	6	0.8	4	US-08-861-464-2	Sequence 2, Appli	655	6	US-08-097-829-4	Sequence 4, Appli
583	6	0.8	4	US-08-396-001-2	Sequence 2, Appli	656	6	US-08-577-403-4	Sequence 4, Appli
584	6	0.8	4	US-09-323-433A-2	Sequence 2, Appli	657	6	US-09-252-991A-31115	Sequence 31115, A
585	6	0.8	4	US-08-311-731A-70	Sequence 70, Appl	658	6	US-09-043-123-4	Sequence 4, Appli
586	6	0.8	4	US-09-134-001C-3935	Sequence 3935, Ap	659	6	US-08-973-398-2	Sequence 2, Appli
587	6	0.8	4	US-09-252-991A-19865	Sequence 19865, A	660	6	PCT-US96-05800-7	Sequence 7, Appli
588	6	0.8	4	US-09-252-991A-20884	Sequence 20884, A	661	6	US-08-090-523-8	Sequence 8, Appli
589	6	0.8	4	US-09-252-991A-28627	Sequence 28627, A	662	6	US-08-398-627-8	Sequence 8, Appli
590	6	0.8	4	US-09-328-352-5272	Sequence 5272, Ap	663	6	US-08-406-858-8	Sequence 8, Appli
591	6	0.8	4	US-08-272-455-14	Sequence 14, Appl	664	6	US-08-120-703A-8	Sequence 8, Appli
592	6	0.8	4	PCT-US95-08565-14	Sequence 14, Appl	665	6	US-08-399-023-8	Sequence 8, Appli
593	6	0.8	4	US-09-134-001C-4663	Sequence 4663, Ap	666	6	PCT-US91-04036-8	Sequence 8, Appli
594	6	0.8	4	US-08-971-782-2	Sequence 2, Appli	667	6	US-09-457-040B-23	Sequence 23, Appl
595	6	0.8	4	US-09-309-026-2	Sequence 2, Appli	668	6	US-09-252-991A-28065	Sequence 8, Appli
596	6	0.8	4	US-09-252-991A-23321	Sequence 23321, A	669	6	US-09-252-991A-17791	Sequence 17791, A
597	6	0.8	4	US-09-134-001C-4107	Sequence 4107, Ap	670	6	US-09-107-532A-5743	Sequence 5743, Ap
598	6	0.8	4	US-09-647-540A-2	Sequence 2, Appli	671	6	US-09-252-991A-26690	Sequence 26690, A
599	6	0.8	4	US-10-119-600-2	Sequence 2, Appli	672	6	US-09-188-930-185	Sequence 185, App
600	6	0.8	4	US-09-252-991A-19007	Sequence 19007, A	673	6	US-09-312-283C-185	Sequence 185, App
601	6	0.8	4	US-09-328-352-5280	Sequence 5280, Ap	674	6	US-09-457-040B-23	Sequence 23, Appl
602	6	0.8	4	US-09-160-036-12	Sequence 12, Appl	675	6	US-09-252-991A-29617	Sequence 29617, A
603	6	0.8	4	US-09-328-352-7117	Sequence 7117, Ap	676	6	US-09-252-991A-30757	Sequence 30757, A
604	6	0.8	4	US-09-252-991A-25780	Sequence 25780, A	677	6	US-09-621-451-2	Sequence 2, Appli
605	6	0.8	4	US-09-252-991A-28932	Sequence 28932, A	678	6	US-09-252-991A-27557	Sequence 27557, A
606	6	0.8	4	US-09-632-636-4	Sequence 4, Appli	679	6	US-09-328-352-4514	Sequence 4514, Ap
607	6	0.8	4	US-08-535-276-4	Sequence 4, Appli	680	6	US-09-529-093A-2	Sequence 2, Appli
608	6	0.8	4	US-09-126-420A-20	Sequence 20, Appl	681	6	US-09-529-154-2	Sequence 12, Appl
609	6	0.8	4	US-09-335-234-4	Sequence 4, Appli	682	6	US-09-934-901-12	Sequence 12, Appl
610	6	0.8	4	US-08-252-991A-32068	Sequence 32068, A	683	6	US-09-252-991A-17217	Sequence 17217, A
611	6	0.8	4	US-08-417-330A-20	Sequence 20, Appl	684	6	US-09-328-352-7955	Sequence 7955, Ap

685	6	0.8	550	4	US-09-396-478A-2	Sequence 2, Appli	758	6	0.8	659	4	US-09-292-858B-21	Sequence 21, Appli
686	6	0.8	553	4	US-09-252-991A-22589	Sequence 22589, A	759	6	0.8	659	4	US-09-252-991A-22940	Sequence 22940, A
687	6	0.8	554	4	US-09-252-991A-23106	Sequence 23106, A	760	6	0.8	659	4	US-08-252-991A-26013	Sequence 26013, A
688	6	0.8	555	1	US-08-039-777-3	Sequence 3, Appli	761	6	0.8	661	1	US-08-394-326-2	Sequence 2, Appli
689	6	0.8	555	1	US-08-611-361A-3	Sequence 3, Appli	762	6	0.8	661	3	US-09-082-306-2	Sequence 2, Appli
690	6	0.8	563	2	US-09-031-392-2	Sequence 2, Appli	763	6	0.8	661	4	US-09-252-991A-18960	Sequence 18960, A
691	6	0.8	563	3	US-09-299-549-2	Sequence 2, Appli	764	6	0.8	671	4	US-09-198-452A-468	Sequence 468, App
692	6	0.8	563	4	US-09-610-417-2	Sequence 2, Appli	765	6	0.8	671	4	US-09-556-916-26	Sequence 26, Appli
693	6	0.8	566	4	US-09-996-243-41	Sequence 41, Appli	766	6	0.8	672	4	US-08-556-916-28	Sequence 28, Appli
694	6	0.8	566	4	US-09-252-991A-22461	Sequence 22461, A	767	6	0.8	672	4	US-09-252-991A-26458	Sequence 26458, A
695	6	0.8	577	3	US-09-315-794-32	Sequence 32, Appli	768	6	0.8	673	4	US-09-252-991A-18107	Sequence 18107, A
696	6	0.8	577	3	US-09-389-341-32	Sequence 32, Appli	769	6	0.8	675	4	US-09-252-991A-23185	Sequence 23185, A
697	6	0.8	581	4	US-08-311-731A-286	Sequence 286, App	770	6	0.8	676	4	US-09-252-991A-19785	Sequence 19785, A
698	6	0.8	582	3	US-09-413-814-83	Sequence 83, Appli	771	6	0.8	680	4	US-09-252-991A-25330	Sequence 25330, A
699	6	0.8	583	4	US-09-328-352-6422	Sequence 422, Ap	772	6	0.8	681	2	US-08-272-255-6	Sequence 6, Appli
700	6	0.8	584	3	US-09-102-528-12	Sequence 12, Appli	773	6	0.8	681	2	US-08-272-255-9	Sequence 9, Appli
701	6	0.8	585	4	US-09-252-991A-32321	Sequence 32321, A	774	6	0.8	681	3	US-08-964-288-5	Sequence 5, Appli
702	6	0.8	585	4	US-09-328-352-6426	Sequence 6426, Ap	775	6	0.8	681	4	US-09-105-254-5	Sequence 5, Appli
703	6	0.8	586	4	US-09-252-991A-28275	Sequence 28275, A	776	6	0.8	681	5	PCT-US95-08565-6	Sequence 6, Appli
704	6	0.8	586	4	US-08-459-346-12	Sequence 12, Appli	777	6	0.8	681	5	PCT-US95-08565-9	Sequence 9, Appli
705	6	0.8	588	3	US-08-889-419-12	Sequence 12, Appli	778	6	0.8	684	4	US-09-252-991A-28604	Sequence 28604, A
706	6	0.8	588	4	US-08-402-542-12	Sequence 42, Appli	779	6	0.8	684	4	US-09-328-352-7195	Sequence 7195, Ap
707	6	0.8	588	4	US-08-714-741-42	Sequence 42, Appli	780	6	0.8	687	4	US-09-328-352-5840	Sequence 5840, Ap
708	6	0.8	588	5	PCT-US93-07189-12	Sequence 12, Appli	781	6	0.8	690	4	US-09-252-991A-23187	Sequence 23187, A
709	6	0.8	590	1	US-08-221-817-14	Sequence 14, Appli	782	6	0.8	698	4	US-09-252-991A-30206	Sequence 30206, A
710	6	0.8	590	1	US-08-454-439-14	Sequence 14, Appli	783	6	0.8	703	4	US-09-252-991A-20065	Sequence 20065, A
711	6	0.8	590	3	US-08-464-954A-5	Sequence 5, Appli	784	6	0.8	703	4	US-09-252-991A-22783	Sequence 22783, A
712	6	0.8	590	4	US-09-252-991A-18898	Sequence 18898, A	785	6	0.8	705	4	US-09-252-991A-26000	Sequence 26000, A
713	6	0.8	590	4	US-09-312-283C-409	Sequence 409, App	786	6	0.8	705	4	US-09-252-991A-27442	Sequence 27442, A
714	6	0.8	590	5	PCT-US94-10487-14	Sequence 14, Appli	787	6	0.8	708	4	US-09-252-991A-18074	Sequence 18074, A
715	6	0.8	592	4	US-09-934-901-14	Sequence 14, Appli	788	6	0.8	715	4	US-09-252-991A-32740	Sequence 32740, A
716	6	0.8	592	4	US-08-836-620A-16	Sequence 16, Appli	789	6	0.8	715	4	US-09-307-143-2	Sequence 2, Appli
717	6	0.8	596	4	US-09-252-991A-23812	Sequence 23812, A	790	6	0.8	724	4	US-09-562-737-23	Sequence 23, Appli
718	6	0.8	600	2	US-08-370-156-4	Sequence 4, Appli	791	6	0.8	724	4	US-09-562-737-23	Sequence 23, Appli
719	6	0.8	600	3	US-08-814-095-4	Sequence 4, Appli	792	6	0.8	728	2	US-09-252-991A-23613	Sequence 23613, A
720	6	0.8	600	3	US-08-975-084-1	Sequence 1, Appli	793	6	0.8	732	4	US-08-533-669A-18	Sequence 18, Appli
721	6	0.8	607	4	US-08-569-804-10	Sequence 10, Appli	794	6	0.8	732	4	US-09-307-143-4	Sequence 4, Appli
722	6	0.8	607	4	US-09-252-991A-20896	Sequence 20896, A	795	6	0.8	732	4	US-09-183-861-18	Sequence 18, Appli
723	6	0.8	607	4	US-09-252-991A-30814	Sequence 30814, A	796	6	0.8	732	4	US-09-022-765-18	Sequence 18, Appli
724	6	0.8	609	3	US-08-980-115-11	Sequence 11, Appli	797	6	0.8	732	4	US-08-671-757A-8	Sequence 8, Appli
725	6	0.8	609	4	US-09-569-804-11	Sequence 11, Appli	798	6	0.8	732	4	US-08-671-757A-8	Sequence 8, Appli
726	6	0.8	613	4	US-09-252-991A-24982	Sequence 24982, A	799	6	0.8	732	4	US-09-551-974A-18	Sequence 18, Appli
727	6	0.8	614	1	US-07-732-962A-2	Sequence 2, Appli	800	6	0.8	740	3	US-08-891-322-2	Sequence 2, Appli
728	6	0.8	614	2	US-08-370-156-2	Sequence 2, Appli	801	6	0.8	740	3	US-09-277-019-2	Sequence 2, Appli
729	6	0.8	614	3	US-08-446-100-19	Sequence 19, Appli	802	6	0.8	748	3	US-08-392-459-24	Sequence 24, Appli
730	6	0.8	614	3	US-08-446-100-20	Sequence 20, Appli	803	6	0.8	748	3	US-08-392-459-28	Sequence 28, Appli
731	6	0.8	614	3	US-08-446-100-21	Sequence 21, Appli	804	6	0.8	748	3	US-08-392-459-34	Sequence 34, Appli
732	6	0.8	614	3	US-08-446-100-22	Sequence 22, Appli	805	6	0.8	748	5	PCT-US91-08525-24	Sequence 24, Appli
733	6	0.8	614	3	US-08-446-100-23	Sequence 23, Appli	806	6	0.8	748	5	PCT-US91-08525-28	Sequence 28, Appli
734	6	0.8	614	3	US-08-446-100-25	Sequence 25, Appli	807	6	0.8	748	5	PCT-US91-08525-34	Sequence 34, Appli
735	6	0.8	614	3	US-08-814-095-2	Sequence 2, Appli	808	6	0.8	748	5	PCT-US93-04384-4	Sequence 4, Appli
736	6	0.8	614	5	PCT-US92-06106-2	Sequence 2, Appli	809	6	0.8	748	5	PCT-US93-04384-6	Sequence 6, Appli
737	6	0.8	615	4	US-09-252-991A-28135	Sequence 28135, A	810	6	0.8	749	4	US-09-562-737-91	Sequence 91, Appli
738	6	0.8	617	2	US-08-370-156-6	Sequence 6, Appli	811	6	0.8	749	4	US-09-562-737-91	Sequence 91, Appli
739	6	0.8	617	3	US-08-814-095-6	Sequence 6, Appli	812	6	0.8	749	4	US-09-562-737-94	Sequence 94, Appli
740	6	0.8	618	3	US-08-834-306-65	Sequence 65, Appli	813	6	0.8	749	4	US-09-562-737-95	Sequence 95, Appli
741	6	0.8	618	3	US-08-993-674A-65	Sequence 65, Appli	814	6	0.8	749	4	US-09-562-737-96	Sequence 96, Appli
742	6	0.8	618	4	US-09-256-976-65	Sequence 65, Appli	815	6	0.8	749	4	US-09-562-737-99	Sequence 99, Appli
743	6	0.8	619	4	US-09-107-532A-4897	Sequence 4897, Ap	816	6	0.8	749	4	US-09-328-352-7588	Sequence 7588, Ap
744	6	0.8	633	4	US-09-328-352-4599	Sequence 4599, Ap	817	6	0.8	752	4	US-09-252-991A-17355	Sequence 17355, A
745	6	0.8	634	4	US-09-252-991A-33132	Sequence 33132, A	818	6	0.8	752	4	US-09-252-991A-17987	Sequence 17987, A
746	6	0.8	643	2	US-08-245-511-47	Sequence 47, Appli	819	6	0.8	753	4	US-09-328-352-5412	Sequence 5412, Ap
747	6	0.8	643	2	US-08-600-993A-47	Sequence 47, Appli	820	6	0.8	757	4	US-09-252-991A-20231	Sequence 20231, A
748	6	0.8	645	4	US-09-252-991A-25253	Sequence 25253, A	821	6	0.8	758	4	US-09-735-101-2	Sequence 2, Appli
749	6	0.8	650	4	US-09-252-991A-20375	Sequence 20375, A	822	6	0.8	765	4	US-09-252-991A-29382	Sequence 29382, A
750	6	0.8	652	4	US-09-252-991A-26354	Sequence 26354, A	823	6	0.8	770	4	US-09-252-991A-24516	Sequence 24516, A
751	6	0.8	652	4	US-09-328-352-5587	Sequence 5587, Ap	824	6	0.8	773	4	US-09-252-991A-23192	Sequence 23192, A
752	6	0.8	653	4	US-09-252-991A-27553	Sequence 27553, A	825	6	0.8	775	4	US-09-328-352-7095	Sequence 7095, Ap
753	6	0.8	655	4	US-09-252-991A-21806	Sequence 21806, A	826	6	0.8	775	4	US-08-817-832B-31	Sequence 31, Appli
754	6	0.8	656	4	US-09-252-991A-29575	Sequence 29575, A	827	6	0.8	782	1	US-07-725-083-2	Sequence 2, Appli
755	6	0.8	657	4	US-09-252-991A-23156	Sequence 23156, A	828	6	0.8	782	3	US-08-669-286-10	Sequence 10, Appli
756	6	0.8	659	1	US-08-198-446B-13	Sequence 13, Appli	829	6	0.8	782	3	US-09-469-253-10	Sequence 10, Appli
757	6	0.8	659	2	US-08-870-693-13	Sequence 13, Appli	830	6	0.8	782	3	US-09-642-146-10	Sequence 10, Appli

831	6	0.8	786	4	US-09-509-802-2	Sequence 2, Appli	904	6	0.8	969	5	PCT-US92-06888-3	Sequence 3, Appli
832	6	0.8	787	3	US-09-188-930-334	Sequence 334, App	905	6	0.8	969	5	PCT-US93-03027-4	Sequence 4, Appli
833	6	0.8	787	4	US-09-312-283C-334	Sequence 334, App	906	6	0.8	970	4	US-09-795-927-7	Sequence 7, Appli
834	6	0.8	790	4	US-09-634-957-2	Sequence 2, Appli	907	6	0.8	976	4	US-08-311-731A-363	Sequence 363, App
835	6	0.8	790	4	US-09-723-153-2	Sequence 2, Appli	908	6	0.8	980	4	US-09-110-517-4	Sequence 4, Appli
836	6	0.8	790	4	US-09-723-429-2	Sequence 2, Appli	909	6	0.8	990	4	US-09-252-991A-32469	Sequence 32469, A
837	6	0.8	792	4	US-09-252-991A-25823	Sequence 25823, A	910	6	0.8	994	4	US-09-521-511C-13	Sequence 13, Appli
838	6	0.8	793	4	US-09-523-849-32	Sequence 32, Appli	911	6	0.8	1007	4	US-09-252-991A-18614	Sequence 18614, A
839	6	0.8	797	3	US-08-433-522A-2	Sequence 4, Appli	912	6	0.8	1008	4	US-09-328-352-7403	Sequence 7403, Ap
840	6	0.8	797	3	US-08-433-522A-4	Sequence 4, Appli	913	6	0.8	1012	2	US-08-475-891A-4	Sequence 4, Appli
841	6	0.8	797	3	US-08-433-522A-6	Sequence 6, Appli	914	6	0.8	1025	2	US-08-567-375-4	Sequence 4, Appli
842	6	0.8	797	3	US-09-135-166-2	Sequence 2, Appli	915	6	0.8	1025	2	US-08-587-680A-4	Sequence 2, Appli
843	6	0.8	797	3	US-09-135-166-4	Sequence 4, Appli	916	6	0.8	1025	5	PCT-US93-04567-2	Sequence 4794, Ap
844	6	0.8	797	3	US-09-135-166-6	Sequence 6, Appli	917	6	0.8	1037	4	US-09-134-001C-4794	Sequence 22493, A
845	6	0.8	797	3	US-08-942-046-2	Sequence 2, Appli	918	6	0.8	1044	4	US-09-252-991A-22493	Sequence 5229, Ap
846	6	0.8	797	3	US-08-942-046-4	Sequence 4, Appli	919	6	0.8	1044	4	US-09-107-532A-5229	Sequence 9, Appli
847	6	0.8	797	3	US-08-942-046-6	Sequence 6, Appli	920	6	0.8	1065	3	US-08-630-172-9	Sequence 9, Appli
848	6	0.8	798	4	US-09-134-001C-4917	Sequence 4917, Ap	921	6	0.8	1065	3	US-09-375-419-9	Sequence 4, Appli
849	6	0.8	799	4	US-09-252-991A-26622	Sequence 26622, A	922	6	0.8	1074	1	US-07-912-015-4	Sequence 20, Appli
850	6	0.8	802	3	US-09-081-345-18	Sequence 18, Appli	923	6	0.8	1082	4	US-09-336-643A-20	Sequence 2, Appli
851	6	0.8	812	4	US-09-252-991A-30710	Sequence 30710, A	924	6	0.8	1083	4	US-09-600-776-2	Sequence 1, Appli
852	6	0.8	816	4	US-09-252-991A-23292	Sequence 23292, A	925	6	0.8	1083	4	US-09-343-494-1	Sequence 2, Appli
853	6	0.8	817	4	US-09-252-991A-23292	Sequence 23292, A	926	6	0.8	1083	4	US-09-358-383C-2	Sequence 2, Appli
854	6	0.8	827	4	US-09-252-991A-27609	Sequence 27609, A	927	6	0.8	1090	4	US-09-252-991A-23149	Sequence 23149, A
855	6	0.8	832	4	US-09-252-991A-20443	Sequence 20443, A	928	6	0.8	1104	4	US-09-252-991A-25463	Sequence 25463, A
856	6	0.8	833	4	US-09-252-991A-22753	Sequence 22753, A	929	6	0.8	1123	4	US-09-252-991A-27976	Sequence 27976, A
857	6	0.8	836	4	US-09-514-302-3	Sequence 3, Appli	930	6	0.8	1132	4	US-09-198-452A-466	Sequence 466, App
858	6	0.8	837	4	US-09-417-197-61	Sequence 61, Appli	931	6	0.8	1138	4	US-09-252-991A-20291	Sequence 20291, A
859	6	0.8	837	4	US-09-122-126B-2	Sequence 2, Appli	932	6	0.8	1140	3	US-09-220-081-2	Sequence 2, Appli
860	6	0.8	842	4	US-09-634-286A-2	Sequence 2, Appli	933	6	0.8	1140	4	US-09-677-575-2	Sequence 2, Appli
861	6	0.8	847	4	US-09-417-197-43	Sequence 43, Appli	934	6	0.8	1148	2	US-08-313-185-58	Sequence 58, Appli
862	6	0.8	857	4	US-09-252-991A-31764	Sequence 31764, A	935	6	0.8	1148	3	US-09-082-614A-58	Sequence 58, Appli
863	6	0.8	858	4	US-09-252-991A-25029	Sequence 25029, A	936	6	0.8	1148	3	US-09-106-075A-90	Sequence 90, Appli
864	6	0.8	861	4	US-09-784-316-2	Sequence 2, Appli	937	6	0.8	1170	2	US-08-789-078-2	Sequence 2, Appli
865	6	0.8	864	4	US-09-252-991A-28455	Sequence 28455, A	938	6	0.8	1170	2	US-08-752-633-2	Sequence 2, Appli
866	6	0.8	871	2	US-08-714-741-40	Sequence 40, Appli	939	6	0.8	1170	2	US-08-476-062A-42	Sequence 42, Appli
867	6	0.8	874	4	US-09-775-009-34	Sequence 34, Appli	940	6	0.8	1170	5	PCT-US95-04886-2	Sequence 2, Appli
868	6	0.8	876	1	US-09-369-364A-15	Sequence 15, Appli	941	6	0.8	1170	5	PCT-US96-01314-42	Sequence 42, Appli
869	6	0.8	876	2	US-08-717-515-4	Sequence 4, Appli	942	6	0.8	1220	2	US-08-680-326-38	Sequence 38, Appli
870	6	0.8	876	4	US-08-510-215A-2	Sequence 2, Appli	943	6	0.8	1231	4	US-08-107-532A-5150	Sequence 5150, Ap
871	6	0.8	876	4	US-09-517-871-20	Sequence 20, Appli	944	6	0.8	1231	4	US-08-930-996A-4	Sequence 4, Appli
872	6	0.8	876	4	US-09-252-991A-22193	Sequence 22193, A	945	6	0.8	1269	3	US-08-781-891-74	Sequence 74, Appli
873	6	0.8	879	4	US-09-252-991A-31974	Sequence 31974, A	946	6	0.8	1269	3	US-08-618-166-74	Sequence 74, Appli
874	6	0.8	885	1	US-09-914-259-38	Sequence 38, Appli	947	6	0.8	1269	3	US-08-618-166-74	Sequence 74, Appli
875	6	0.8	885	1	US-08-484-105-14	Sequence 14, Appli	948	6	0.8	1277	4	US-08-717-515-8	Sequence 8, Appli
876	6	0.8	881	4	US-08-484-106-14	Sequence 14, Appli	949	6	0.8	1277	4	US-09-397-885-3	Sequence 3, Appli
877	6	0.8	882	4	US-09-585-858-4	Sequence 4, Appli	950	6	0.8	1277	4	US-09-969-362-3	Sequence 3, Appli
878	6	0.8	882	4	US-09-585-858-2	Sequence 2, Appli	951	6	0.8	1328	3	US-08-781-891-76	Sequence 76, Appli
879	6	0.8	882	4	US-09-585-858-3	Sequence 3, Appli	952	6	0.8	1328	3	US-08-618-166-76	Sequence 16, Appli
880	6	0.8	882	4	US-09-252-991A-32194	Sequence 32194, A	953	6	0.8	1375	3	US-09-171-410-1	Sequence 1, Appli
881	6	0.8	901	2	US-08-884-681-5	Sequence 5, Appli	954	6	0.8	1415	4	US-09-252-991A-26438	Sequence 26438, A
882	6	0.8	906	3	US-09-258-643-5	Sequence 5, Appli	955	6	0.8	1415	4	US-09-252-991A-31822	Sequence 31822, A
883	6	0.8	906	3	US-08-190-802A-31	Sequence 31, Appli	956	6	0.8	1423	3	US-08-810-712-10	Sequence 10, Appli
884	6	0.8	906	3	US-08-477-346-31	Sequence 31, Appli	957	6	0.8	1430	4	US-09-252-991A-18190	Sequence 18190, A
885	6	0.8	906	4	US-08-473-089-31	Sequence 31, Appli	958	6	0.8	1451	1	US-08-308-872B-4	Sequence 4, Appli
886	6	0.8	911	3	US-08-487-072A-31	Sequence 31, Appli	959	6	0.8	1454	3	US-08-392-459-22	Sequence 22, Appli
887	6	0.8	915	4	US-08-460-269C-4	Sequence 4, Appli	960	6	0.8	1454	3	US-08-392-459-26	Sequence 26, Appli
888	6	0.8	922	3	US-09-252-991A-24992	Sequence 24992, A	961	6	0.8	1454	5	PCT-US91-08525-22	Sequence 22, Appli
889	6	0.8	923	4	US-08-460-269C-6	Sequence 6, Appli	962	6	0.8	1454	5	PCT-US91-08525-26	Sequence 26, Appli
890	6	0.8	923	4	US-09-397-885-1	Sequence 1, Appli	963	6	0.8	1454	5	PCT-US91-08525-32	Sequence 32, Appli
891	6	0.8	934	1	US-09-969-362-1	Sequence 1, Appli	964	6	0.8	1454	5	PCT-US93-04384-2	Sequence 2, Appli
892	6	0.8	934	3	US-08-925-805A-80	Sequence 80, Appli	965	6	0.8	1454	5	PCT-US93-04384-8	Sequence 8, Appli
893	6	0.8	943	3	US-08-929-329-6	Sequence 6, Appli	966	6	0.8	1454	5	PCT-US93-04384-12	Sequence 12, Appli
894	6	0.8	943	4	US-09-397-885-5	Sequence 5, Appli	967	6	0.8	1454	5	PCT-US93-04384-16	Sequence 16, Appli
895	6	0.8	943	4	US-09-969-362-5	Sequence 5, Appli	968	6	0.8	1454	5	PCT-US93-04384-43	Sequence 43, Appli
896	6	0.8	945	4	US-09-198-452A-112	Sequence 112, App	969	6	0.8	1454	5	PCT-US93-04384-44	Sequence 44, Appli
897	6	0.8	956	4	US-09-252-991A-17124	Sequence 17124, A	970	6	0.8	1454	5	PCT-US93-04384-45	Sequence 45, Appli
898	6	0.8	966	4	US-09-207-359B-47	Sequence 47, Appli	971	6	0.8	1454	5	PCT-US93-04384-46	Sequence 46, Appli
899	6	0.8	968	3	US-08-560-005-7	Sequence 7, Appli	972	6	0.8	1454	5	PCT-US93-04384-47	Sequence 47, Appli
900	6	0.8	968	3	US-09-418-540-7	Sequence 7, Appli	973	6	0.8	1454	5	PCT-US93-04384-48	Sequence 48, Appli
901	6	0.8	968	4	US-09-969-328-3	Sequence 3, Appli	974	6	0.8	1477	1	US-08-038-682-4	Sequence 4, Appli
902	6	0.8	969	1	US-08-365-689-3	Sequence 3, Appli	975	6	0.8	1477	1	US-08-302-832-4	Sequence 4, Appli
903	6	0.8	969	1	US-07-747-781-3	Sequence 3, Appli	976	6	0.8	1477	2	US-08-530-198-4	Sequence 4, Appli

977 6 0.8 1477 2 US-08-469-880-4 Sequence 4, Appli
978 6 0.8 1477 2 US-08-728-470-4 Sequence 4, Appli
979 6 0.8 1477 3 US-08-617-697-4 Sequence 4, Appli
980 6 0.8 1477 3 US-08-492-459-10 Sequence 10, Appli
981 6 0.8 1477 3 US-08-423-752-10 Sequence 10, Appli
982 6 0.8 1477 3 US-08-945-994-3 Sequence 3, Appli
983 6 0.8 1477 3 US-08-716-873-24 Sequence 24, Appli
984 6 0.8 1477 3 US-08-719-641-4 Sequence 4, Appli
985 6 0.8 1477 3 US-09-368-431-24 Sequence 24, Appli
986 6 0.8 1477 4 US-09-414-006-10 Sequence 10, Appli
987 6 0.8 1477 4 US-09-447-223-10 Sequence 10, Appli
988 6 0.8 1477 4 US-09-206-942-71 Sequence 71, Appli
989 6 0.8 1507 3 US-08-929-329-5 Sequence 5, Appli
990 6 0.8 1507 6 5268270-2 Patent No. 5268270
991 6 0.8 1514 2 US-08-853-310-4 Sequence 4, Appli
992 6 0.8 1536 1 US-08-038-682-2 Sequence 2, Appli
993 6 0.8 1536 1 US-08-302-832-2 Sequence 2, Appli
994 6 0.8 1536 2 US-08-530-198-2 Sequence 2, Appli
995 6 0.8 1536 2 US-08-469-880-2 Sequence 2, Appli
996 6 0.8 1536 2 US-08-728-470-2 Sequence 2, Appli
997 6 0.8 1536 2 US-08-617-697-2 Sequence 2, Appli
998 6 0.8 1536 3 US-08-719-641-2 Sequence 2, Appli
999 6 0.8 1536 4 US-09-206-942-67 Sequence 67, Appli
1000 6 0.8 1580 2 US-08-804-227C-11 Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-2981
; Sequence 2981, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2981
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2981

Query Match 1.1%; Score 8; DB 4; Length 430;
Best Local Similarity 100.0%; Pred. No. 14; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 9 IVLSILLI 16
Db 231 IVLSILLI 238

RESULT 2
US-09-199-637A-267
; Sequence 267, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John

; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 80786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIOR FILING DATE: 1998-11-25
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 267
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-267

Query Match 1.1%; Score 8; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 15; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 238 HRAVLGWR 245
Db 65 HRAVLGWR 72

RESULT 3
US-09-328-352-5144
; Sequence 5144, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5144
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5144

Query Match 1.1%; Score 8; DB 4; Length 704;
Best Local Similarity 100.0%; Pred. No. 23; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 256 RRDQYGLP 263
Db 267 RRDQYGLP 274

RESULT 4
US-09-162-366C-18
; Sequence 18, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18

```
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-18

Query Match          0.9%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 1 NLGDALD 7

RESULT 5
US-09-162-366C-17
; Sequence 17, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-17

Query Match          0.9%; Score 7; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 2 NLGDALD 8

RESULT 6
US-09-162-366C-16
; Sequence 16, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
```

```
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-16

Query Match          0.9%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 3 NLGDALD 9

RESULT 7
US-09-162-366C-1
; Sequence 1, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-1

Query Match          0.9%; Score 7; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 4 NLGDALD 10

RESULT 8
US-09-162-366C-4
; Sequence 4, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
```

```

; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-4

```

```

Query Match 0.9%; Score 7; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 5 NLGDALD 11

```

```

RESULT 9
US-09-162-366C-14
; Sequence 14, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-14

```

```

Query Match 0.9%; Score 7; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 5 NLGDALD 11

```

```

RESULT 10
US-09-162-366C-12
; Sequence 12, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil

```

```

; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-12

```

```

Query Match 0.9%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 6 NLGDALD 12

```

```

RESULT 11
US-09-162-366C-13
; Sequence 13, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-13

```

```

Query Match 0.9%; Score 7; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 7 NLGDALD 13

```

```

RESULT 12
US-09-162-366C-11

```

; Sequence 11, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-11

Query Match 0.9%; Score 7; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14
|||||

RESULT 13
US-09-162-366C-3
; Sequence 3, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-3

Query Match 0.9%; Score 7; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14
|||||

RESULT 14
US-09-162-366C-2
; Sequence 2, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-2

Query Match 0.9%; Score 7; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 7 NLGDALD 13
|||||

RESULT 15
US-09-205-258-963
; Sequence 963, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894

DB 14 AGLYDVA 20

RESULT 16

US-08-928-361B-10

Sequence 10, Application US/08928361B

Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn

TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS, THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

TITLE OF INVENTION: SPECIES INFECTIONS

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: PETERS, VERNY, JONES & BIKSA

STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/928,361B

FILING DATE: 12-SEP-1997

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/026,062

FILING DATE: 13-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Verity, Hana

REGISTRATION NUMBER: 30,518

REFERENCE/DOCKET NUMBER: 480.76-1 (HV)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-1677

TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 112 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-928-361B-10

Query Match 0.9%; Score 7; DB 3; Length 112;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLKPIV 10

DB 74 TTLKPIV 80

RESULT 17

US-08-928-361B-29

Sequence 29, Application US/08928361B

Patent No. 6071518

GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn

TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS, THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS

TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

TITLE OF INVENTION: SPECIES INFECTIONS

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: PETERS, VERNY, JONES & BIKSA

STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,971

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,964

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,882

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,899

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,893

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,900

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,901

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,892

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,915

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,019

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,970

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,972

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,916

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/049,373

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,875

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/049,374

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,917

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,949

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,974

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,883

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,897

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,898

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,962

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,963

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,877

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,878

EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/070,923

EARLIER FILING DATE: 1997-12-18

EARLIER APPLICATION NUMBER: 60/092,921

EARLIER FILING DATE: 1998-07-15

EARLIER APPLICATION NUMBER: 60/094,657

EARLIER FILING DATE: 1998-07-30

NUMBER OF SEQ ID NOS: 1227

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 963

LENGTH: 51

TYPE: PRT

ORGANISM: Homo sapiens

US-09-205-258-963

Query Match 0.9%; Score 7; DB 4; Length 51;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLYDVA 159

|||||||

STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verity, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 112 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-29

Query Match 0.9%; Score 7; DB 3; Length 112;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLKPIV 10
Db 74 TTLKPIV 80

RESULT 18

US-09-588-995A-10
Sequence 10, Application US/09588995A
Patent No. 6514697
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: BARNES, DEBRA A.
APPLICANT: NELSON, RICHARD C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: INFECTIONS
FILE REFERENCE: 480.19-5
CURRENT APPLICATION NUMBER: US/09/588,995A
CURRENT FILING DATE: 2000-06-06
PRIOR APPLICATION NUMBER: 08/827,171
PRIOR FILING DATE: 1997-03-27
PRIOR APPLICATION NUMBER: 08/928,361
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: 08/700,651
PRIOR FILING DATE: 1996-08-14
PRIOR APPLICATION NUMBER: 08/415,751
PRIOR FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 115
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 112
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-09-588-995A-10

Query Match 0.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 TTLKPIV 10
Db 74 TTLKPIV 80

RESULT 19

US-09-252-991A-23722
Sequence 23722, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23722
LENGTH: 141
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (86)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-23722

Query Match 0.9%; Score 7; DB 4; Length 141;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 APRVPA 662
Db 13 APRVPA 19

RESULT 20

US-09-252-991A-30346
Sequence 30346, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30346
LENGTH: 167
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30346

Query Match 0.9%; Score 7; DB 4; Length 167;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 PRVPAAR 663
Db 112 PRVPAAR 118

RESULT 21
US-09-252-991A-20206
; Sequence 20206, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20206
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20206

Query Match 0.9%; Score 7; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 SSALSAT 412
Db 138 SSALSAT 144
|||||

RESULT 22
US-08-415-751-1
; Sequence 1, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS AND KIT
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPPIO & FINLEY
; STREET: 385 Shekman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (IHD)
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Cryptosporidium parvum
US-08-415-751-1

Query Match 0.9%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLXPV 10
Db 69 TTLXPV 75
|||||

RESULT 23
US-09-702-114A-15
; Sequence 15, Application US/09702114A
; Patent No. 6566078
; GENERAL INFORMATION:
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; APPLICANT: Mary Faris
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: 36P6D5: SECRETED TUMOR ANTIGEN
; FILE REFERENCE: 129,22-US-U1
; CURRENT APPLICATION NUMBER: US/09/702,114A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/162,417
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-702-114A-15

Query Match 0.9%; Score 7; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 NGVSGEL 176
Db 67 NGVSGEL 73
|||||

RESULT 24
US-09-252-991A-25202
; Sequence 25202, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25202


```

; LENGTH: 218
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25202

Query Match      0.9%; Score 7; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 212 SGDYAVP 218
Db 125 SGDYAVP 131

RESULT 25
US-09-252-991A-22090
; Sequence 22090, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22090
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22090

Query Match      0.9%; Score 7; DB 4; Length 230;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 ARIELRH 387
Db 81 ARIELRH 87

RESULT 26
US-09-252-991A-30274
; Sequence 30274, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30274
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30274

Query Match      0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 QHRAVLG 243
Db 237 QHRAVLG 243

```

```

Db 115 QHRAVLG 121

RESULT 27
US-09-489-847-205
; Sequence 205, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-489-847-205

Query Match      0.9%; Score 7; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 AGLVDVA 159
Db 206 AGLVDVA 212

RESULT 28
US-09-252-991A-29972
; Sequence 29972, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29972
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29972

Query Match      0.9%; Score 7; DB 4; Length 247;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 GGGASAP 95
Db 61 GGGASAP 67

```

```

RESULT 29
US-09-134-001C-3244
; Sequence 3244, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3244
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3244

Query Match      0.9%; Score 7; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      748 RSFTGGV 754
Db      236 RSFTGGV 242

RESULT 30
US-07-857-224B-22
; Sequence 22, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION:
; ORIGINAL SOURCE:
; ORGANISM: rabbit
; FEATURE: Protein kinase; Table 8 Column 24
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.

```

```

;
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-22

Query Match      0.9%; Score 7; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      489 LSGNWF 495
Db      213 LSGNWF 219

RESULT 31
US-09-205-258-958
; Sequence 958, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972

```

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 958
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-203-258-958

```

```

Query Match      0.9%; Score 7; DB 4; Length 276;
Best Local Similarity 100.0%; Pred.No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 153 AGLVDVA 159
DB 239 AGLVDVA 245

```

```

RESULT 32
US-09-532A-3799
; Sequence 3799, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3799:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...291
; SEQUENCE DESCRIPTION: SEQ ID NO: 3799:
US-09-107-532A-3799

```

```

Query Match      0.9%; Score 7; DB 4; Length 291;
Best Local Similarity 100.0%; Pred.No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 13 ILLINTP 19
DB 240 ILLINTP 246

```

```

RESULT 33
US-09-252-991A-24070
; Sequence 24070, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24070
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24070

```

```

Query Match      0.9%; Score 7; DB 4; Length 308;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 364 EKAGDAV 370
DB 280 EKAGDAV 286

```

```
RESULT 34
US-09-134-001C-2983
; Sequence 2983, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2983
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2983

Query Match      0.9%; Score 7; DB 4; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      73 LGDALDG 79
Db      300 LGDALDG 306

RESULT 35
US-09-252-991A-25015
; Sequence 25015, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25015
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25015

Query Match      0.9%; Score 7; DB 4; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      476 DLGAHRQ 482
Db      138 DLGAHRQ 144

RESULT 36
US-09-134-001C-2890
; Sequence 2890, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
```

```
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2890
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2890

Query Match      0.9%; Score 7; DB 4; Length 321;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      751 TGGVNVK 757
Db      293 TGGVNVK 299

RESULT 37
US-08-415-751-5
; Sequence 5, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS AND KIT
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPLO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; City: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
```

;
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; NAME/KEY: Positions coded by nonsense codons are
; NAME/KEY: identified as Xaa.
US-08-415-751-5

Query Match 0.9%; Score 7; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTKKPIV 10
| | | | |
Db 151 TTKKPIV 157

RESULT 38
US-09-252-991A-25329
; Sequence 25329, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25329
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25329

Query Match 0.9%; Score 7; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 ALDGVPG 82
| | | | |
Db 184 ALDGVPG 190

RESULT 39
US-08-924-183-2
; Sequence 2, Application US/08924183A
; Patent No. 6218109
; GENERAL INFORMATION:
; APPLICANT: Ellledge, Stephen J.
; APPLICANT: Sanchez, Yolanda
; TITLE OF INVENTION: MAMMALIAN CHECKPOINT GENES AND PROTEINS
; FILE REFERENCE: 120541-1003
; CURRENT APPLICATION NUMBER: US/08/924,183A
; CURRENT FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-924-183-2

Query Match 0.9%; Score 7; DB 3; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 RPRATSG 49
| | | | |
Db 275 RPRATSG 281

RESULT 40
US-09-488-364-2
; Sequence 2, Application US/09488364
; Patent No. 6307015
; GENERAL INFORMATION:
; APPLICANT: Ellledge, Stephen J.
; APPLICANT: Sanchez, Yolanda
; TITLE OF INVENTION: MAMMALIAN CHECKPOINT GENES AND PROTEINS
; FILE REFERENCE: 120541-1013
; CURRENT APPLICATION NUMBER: US/09/488,364
; CURRENT FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-488-364-2

Query Match 0.9%; Score 7; DB 4; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 RPRATSG 49
| | | | |
Db 275 RPRATSG 281

Search completed: November 14, 2003, 11:09:15
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:08:49 ; Search time 35 Seconds
(without alignments)
3953.717 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTLKPIVLISILLINPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	23.1	764	15 US-10-065-551-2	Sequence 2, Appli
2	8	1.1	313	15 US-10-204-887-121	Sequence 121, App
3	8	1.1	387	15 US-10-156-761-9479	Sequence 9479, Ap
4	8	1.1	401	12 US-09-996-008B-2	Sequence 2, Appli
5	8	1.1	401	12 US-10-286-326-4	Sequence 4, Appli
6	8	1.1	446	11 US-09-975-719-267	Sequence 267, App
7	7	0.9	7	US-09-915-306-18	Sequence 18, Appl
8	7	0.9	7	US-09-915-374-18	Sequence 17, Appl
9	7	0.9	8	US-09-915-306-17	Sequence 17, Appl
10	7	0.9	8	US-09-915-374-17	Sequence 17, Appl
11	7	0.9	9	US-09-915-306-15	Sequence 16, Appl
12	7	0.9	9	US-09-915-374-16	Sequence 16, Appl
13	7	0.9	10	US-09-915-306-1	Sequence 1, Appli
14	7	0.9	10	US-09-915-374-1	Sequence 1, Appli
15	7	0.9	11	9 US-09-915-306-4	Sequence 4, Appli

16	7	0.9	11	9 US-09-915-306-14	Sequence 14, Appl
17	7	0.9	11	10 US-09-915-374-4	Sequence 4, Appli
18	7	0.9	11	10 US-09-915-374-14	Sequence 14, Appl
19	7	0.9	12	9 US-09-915-306-12	Sequence 12, Appl
20	7	0.9	12	10 US-09-915-374-12	Sequence 12, Appl
21	7	0.9	13	9 US-09-915-306-13	Sequence 13, Appl
22	7	0.9	13	10 US-09-915-374-13	Sequence 13, Appl
23	7	0.9	15	9 US-09-915-306-11	Sequence 11, Appl
24	7	0.9	15	10 US-09-915-374-11	Sequence 11, Appl
25	7	0.9	18	9 US-09-915-306-3	Sequence 3, Appli
26	7	0.9	18	10 US-09-915-374-3	Sequence 3, Appli
27	7	0.9	20	9 US-09-915-306-2	Sequence 2, Appli
28	7	0.9	20	10 US-09-915-374-2	Sequence 2, Appli
29	7	0.9	46	12 US-10-124-805-518	Sequence 518, App
30	7	0.9	46	14 US-10-007-805-518	Sequence 518, App
31	7	0.9	46	15 US-10-076-622-518	Sequence 518, App
32	7	0.9	47	9 US-09-864-761-42772	Sequence 42772, A
33	7	0.9	51	12 US-09-933-767-963	Sequence 963, App
34	7	0.9	51	15 US-10-023-282-963	Sequence 963, App
35	7	0.9	52	15 US-10-156-761-15079	Sequence 15079, A
36	7	0.9	111	15 US-10-156-761-7973	Sequence 7973, Ap
37	7	0.9	127	10 US-09-764-868-801	Sequence 801, App
38	7	0.9	136	12 US-10-123-058-6	Sequence 6, Appli
39	7	0.9	136	12 US-10-414-609-6	Sequence 6, Appli
40	7	0.9	147	15 US-10-153-668-364	Sequence 364, App
41	7	0.9	159	15 US-10-156-761-13330	Sequence 13330, A
42	7	0.9	210	15 US-10-037-340-133	Sequence 133, App
43	7	0.9	230	9 US-09-759-468-2	Sequence 2, Appli
44	7	0.9	232	10 US-09-919-497-78	Sequence 78, Appl
45	7	0.9	234	9 US-09-815-242-5314	Sequence 5314, Ap
46	7	0.9	241	9 US-09-815-242-12602	Sequence 12602, A
47	7	0.9	243	10 US-09-938-418-7	Sequence 7, Appli
48	7	0.9	243	11 US-09-946-374-352	Sequence 352, App
49	7	0.9	243	11 US-09-946-374-431	Sequence 431, App
50	7	0.9	243	12 US-10-015-387A-352	Sequence 352, App
51	7	0.9	243	12 US-10-063-735-122	Sequence 122, App
52	7	0.9	243	12 US-10-006-130A-352	Sequence 352, App
53	7	0.9	243	12 US-10-006-130A-431	Sequence 431, App
54	7	0.9	243	12 US-10-199-672-366	Sequence 366, App
55	7	0.9	243	12 US-10-006-172A-352	Sequence 352, App
56	7	0.9	243	12 US-10-006-172A-431	Sequence 431, App
57	7	0.9	243	12 US-10-187-749-366	Sequence 366, App
58	7	0.9	243	12 US-10-194-457-366	Sequence 366, App
59	7	0.9	243	12 US-10-184-642-366	Sequence 366, App
60	7	0.9	243	12 US-10-196-747-366	Sequence 366, App
61	7	0.9	243	12 US-10-015-392A-352	Sequence 352, App
62	7	0.9	243	12 US-10-015-392A-431	Sequence 431, App
63	7	0.9	243	12 US-10-017-253A-352	Sequence 352, App
64	7	0.9	243	12 US-10-017-253A-431	Sequence 431, App
65	7	0.9	243	12 US-10-124-805-514	Sequence 514, App
66	7	0.9	243	12 US-10-173-689-366	Sequence 366, App
67	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
68	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
69	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
70	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
71	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
72	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
73	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
74	7	0.9	243	12 US-10-173-699-366	Sequence 366, App
75	7	0.9	243	12 US-10-174-569-366	Sequence 366, App
76	7	0.9	243	12 US-10-174-583-366	Sequence 366, App
77	7	0.9	243	12 US-10-174-587-366	Sequence 366, App
78	7	0.9	243	12 US-10-174-587-366	Sequence 366, App
79	7	0.9	243	12 US-10-174-587-366	Sequence 366, App
80	7	0.9	243	12 US-10-175-736-366	Sequence 366, App
81	7	0.9	243	12 US-10-175-736-366	Sequence 366, App
82	7	0.9	243	12 US-10-175-742-366	Sequence 366, App
83	7	0.9	243	12 US-10-175-742-366	Sequence 366, App
84	7	0.9	243	12 US-10-175-748-366	Sequence 366, App
85	7	0.9	243	12 US-10-175-751-366	Sequence 366, App
86	7	0.9	243	12 US-10-175-751-366	Sequence 366, App
87	7	0.9	243	12 US-10-176-480-366	Sequence 366, App
88	7	0.9	243	12 US-10-176-489-366	Sequence 366, App

235	7	0.9	243	12	US-10-013-912A-431	Sequence 431, App	308	7	0.9	243	15	US-10-180-557-366	Sequence 366, App
236	7	0.9	243	12	US-10-015-610A-352	Sequence 352, App	309	7	0.9	243	15	US-10-076-622-514	Sequence 514, App
237	7	0.9	243	12	US-10-015-610A-431	Sequence 431, App	310	7	0.9	243	15	US-10-063-502-122	Sequence 122, App
238	7	0.9	243	12	US-10-015-653A-352	Sequence 352, App	311	7	0.9	243	15	US-10-173-700-366	Sequence 366, App
239	7	0.9	243	12	US-10-015-653A-431	Sequence 431, App	312	7	0.9	243	15	US-10-174-572-366	Sequence 366, App
240	7	0.9	243	12	US-10-015-671A-352	Sequence 352, App	313	7	0.9	243	15	US-10-174-579-366	Sequence 366, App
241	7	0.9	243	12	US-10-015-671A-431	Sequence 431, App	314	7	0.9	243	15	US-10-174-582-366	Sequence 366, App
242	7	0.9	243	12	US-10-063-520-122	Sequence 122, App	315	7	0.9	243	15	US-10-174-588-366	Sequence 366, App
243	7	0.9	243	12	US-10-063-548-122	Sequence 122, App	316	7	0.9	243	15	US-10-175-739-366	Sequence 366, App
244	7	0.9	243	12	US-10-063-578-122	Sequence 122, App	317	7	0.9	243	15	US-10-175-740-366	Sequence 366, App
245	7	0.9	243	12	US-10-063-647-122	Sequence 122, App	318	7	0.9	243	15	US-10-175-743-366	Sequence 366, App
246	7	0.9	243	12	US-10-063-648-122	Sequence 122, App	319	7	0.9	243	15	US-10-176-488-366	Sequence 366, App
247	7	0.9	243	12	US-10-063-677-122	Sequence 122, App	320	7	0.9	243	15	US-10-176-492-366	Sequence 366, App
248	7	0.9	243	12	US-10-063-684-122	Sequence 122, App	321	7	0.9	243	15	US-10-176-747-366	Sequence 366, App
249	7	0.9	243	12	US-10-063-688-122	Sequence 122, App	322	7	0.9	243	15	US-10-176-750-366	Sequence 366, App
250	7	0.9	243	12	US-10-223-081-246	Sequence 246, App	323	7	0.9	243	15	US-10-176-985-366	Sequence 366, App
251	7	0.9	243	12	US-10-012-237A-352	Sequence 352, App	324	7	0.9	243	15	US-10-176-987-366	Sequence 366, App
252	7	0.9	243	12	US-10-012-237A-431	Sequence 431, App	325	7	0.9	243	15	US-10-176-992-366	Sequence 366, App
253	7	0.9	243	12	US-10-013-906A-352	Sequence 352, App	326	7	0.9	243	15	US-10-176-993-366	Sequence 366, App
254	7	0.9	243	12	US-10-013-906A-431	Sequence 431, App	327	7	0.9	243	15	US-10-184-658-366	Sequence 366, App
255	7	0.9	243	12	US-10-015-388A-352	Sequence 352, App	328	7	0.9	243	15	US-10-176-991-366	Sequence 366, App
256	7	0.9	243	12	US-10-015-388A-431	Sequence 431, App	329	7	0.9	243	15	US-10-173-695-366	Sequence 366, App
257	7	0.9	243	12	US-10-015-480A-352	Sequence 352, App	330	7	0.9	243	15	US-10-173-697-366	Sequence 366, App
258	7	0.9	243	12	US-10-015-480A-431	Sequence 431, App	331	7	0.9	243	15	US-10-173-705-366	Sequence 366, App
259	7	0.9	243	12	US-10-015-715A-352	Sequence 352, App	332	7	0.9	243	15	US-10-174-576-366	Sequence 366, App
260	7	0.9	243	12	US-10-015-715A-431	Sequence 431, App	333	7	0.9	243	15	US-10-174-585-366	Sequence 366, App
261	7	0.9	243	12	US-10-063-561-122	Sequence 122, App	334	7	0.9	243	15	US-10-174-586-366	Sequence 366, App
262	7	0.9	243	12	US-10-063-617-122	Sequence 122, App	335	7	0.9	243	15	US-10-175-747-366	Sequence 366, App
263	7	0.9	243	12	US-10-063-618-122	Sequence 122, App	336	7	0.9	243	15	US-10-175-748-366	Sequence 366, App
264	7	0.9	243	12	US-10-063-657-122	Sequence 122, App	337	7	0.9	243	15	US-10-176-481-366	Sequence 366, App
265	7	0.9	243	12	US-10-063-664-122	Sequence 122, App	338	7	0.9	243	15	US-10-176-485-366	Sequence 366, App
266	7	0.9	243	12	US-10-063-668-122	Sequence 122, App	339	7	0.9	243	15	US-10-176-487-366	Sequence 366, App
267	7	0.9	243	12	US-10-063-718-122	Sequence 122, App	340	7	0.9	243	15	US-10-176-493-366	Sequence 366, App
268	7	0.9	243	12	US-10-063-741-122	Sequence 122, App	341	7	0.9	243	15	US-10-176-511-366	Sequence 366, App
269	7	0.9	243	12	US-10-223-082-246	Sequence 246, App	342	7	0.9	243	15	US-10-176-919-366	Sequence 366, App
270	7	0.9	243	12	US-10-012-753A-352	Sequence 352, App	343	7	0.9	243	15	US-10-176-925-366	Sequence 366, App
271	7	0.9	243	12	US-10-012-753A-431	Sequence 431, App	344	7	0.9	243	15	US-10-176-978-366	Sequence 366, App
272	7	0.9	243	12	US-10-015-385A-352	Sequence 352, App	345	7	0.9	243	15	US-10-179-510-366	Sequence 366, App
273	7	0.9	243	12	US-10-015-385A-431	Sequence 431, App	346	7	0.9	243	15	US-10-180-543-366	Sequence 366, App
274	7	0.9	243	12	US-10-007-236A-352	Sequence 352, App	347	7	0.9	243	15	US-10-180-544-366	Sequence 366, App
275	7	0.9	243	12	US-10-007-236A-431	Sequence 431, App	348	7	0.9	243	15	US-10-180-546-366	Sequence 366, App
276	7	0.9	243	12	US-10-015-389A-352	Sequence 352, App	349	7	0.9	243	15	US-10-180-547-366	Sequence 366, App
277	7	0.9	243	12	US-10-015-389A-431	Sequence 431, App	350	7	0.9	243	15	US-10-180-549-366	Sequence 366, App
278	7	0.9	243	12	US-10-013-915A-352	Sequence 352, App	351	7	0.9	243	15	US-10-180-555-366	Sequence 366, App
279	7	0.9	243	12	US-10-013-915A-431	Sequence 431, App	352	7	0.9	243	15	US-10-180-559-366	Sequence 366, App
280	7	0.9	243	12	US-10-015-394A-352	Sequence 352, App	353	7	0.9	243	15	US-10-181-000-366	Sequence 366, App
281	7	0.9	243	12	US-10-015-394A-431	Sequence 431, App	354	7	0.9	243	15	US-10-183-010-366	Sequence 366, App
282	7	0.9	243	12	US-10-015-519A-352	Sequence 352, App	355	7	0.9	243	15	US-10-183-012-366	Sequence 366, App
283	7	0.9	243	12	US-10-015-519A-431	Sequence 431, App	356	7	0.9	243	15	US-10-184-614-366	Sequence 366, App
284	7	0.9	243	12	US-10-063-550-122	Sequence 122, App	357	7	0.9	243	15	US-10-184-623-366	Sequence 366, App
285	7	0.9	243	12	US-10-179-509-366	Sequence 366, App	358	7	0.9	243	15	US-10-184-635-366	Sequence 366, App
286	7	0.9	243	12	US-10-194-486-366	Sequence 366, App	359	7	0.9	243	15	US-10-184-637-366	Sequence 366, App
287	7	0.9	243	12	US-10-195-887-366	Sequence 366, App	360	7	0.9	243	15	US-10-184-646-366	Sequence 366, App
288	7	0.9	243	12	US-10-195-893-366	Sequence 366, App	361	7	0.9	243	15	US-10-184-647-366	Sequence 366, App
289	7	0.9	243	12	US-10-195-900-366	Sequence 366, App	362	7	0.9	243	15	US-10-184-652-366	Sequence 366, App
290	7	0.9	243	12	US-10-198-759-366	Sequence 366, App	363	7	0.9	243	15	US-10-187-594-366	Sequence 366, App
291	7	0.9	243	12	US-10-205-506-366	Sequence 366, App	364	7	0.9	243	15	US-10-187-596-366	Sequence 366, App
292	7	0.9	243	14	US-10-006-867-122	Sequence 122, App	365	7	0.9	243	15	US-10-187-745-366	Sequence 366, App
293	7	0.9	243	14	US-10-052-586-366	Sequence 366, App	366	7	0.9	243	15	US-10-187-885-366	Sequence 366, App
294	7	0.9	243	14	US-10-007-805-514	Sequence 514, App	367	7	0.9	243	15	US-10-187-886-366	Sequence 366, App
295	7	0.9	243	14	US-10-045-992-4	Sequence 4, Appl1	368	7	0.9	243	15	US-10-199-484-366	Sequence 366, App
296	7	0.9	243	14	US-10-063-547-122	Sequence 122, App	369	7	0.9	243	15	US-10-196-756-366	Sequence 366, App
297	7	0.9	243	15	US-10-174-590-366	Sequence 366, App	370	7	0.9	243	15	US-10-176-751-366	Sequence 366, App
298	7	0.9	243	15	US-10-176-758-366	Sequence 366, App	371	7	0.9	243	15	US-10-176-760-366	Sequence 366, App
299	7	0.9	243	15	US-10-175-737-366	Sequence 366, App	372	7	0.9	243	15	US-10-176-990-366	Sequence 366, App
300	7	0.9	243	15	US-10-063-616-122	Sequence 122, App	373	7	0.9	243	15	US-10-180-541-366	Sequence 366, App
301	7	0.9	243	15	US-10-173-706-366	Sequence 366, App	374	7	0.9	243	15	US-10-180-542-366	Sequence 366, App
302	7	0.9	243	15	US-10-175-738-366	Sequence 366, App	375	7	0.9	243	15	US-10-180-548-366	Sequence 366, App
303	7	0.9	243	15	US-10-175-752-366	Sequence 366, App	376	7	0.9	243	15	US-10-180-551-366	Sequence 366, App
304	7	0.9	243	15	US-10-176-482-366	Sequence 366, App	377	7	0.9	243	15	US-10-180-998-366	Sequence 366, App
305	7	0.9	243	15	US-10-176-757-366	Sequence 366, App	378	7	0.9	243	15	US-10-180-999-366	Sequence 366, App
306	7	0.9	243	15	US-10-176-913-366	Sequence 366, App	379	7	0.9	243	15	US-10-183-013-366	Sequence 366, App
307	7	0.9	243	15	US-10-180-552-366	Sequence 366, App	380	7	0.9	243	15	US-10-184-612-366	Sequence 366, App

673	7	0.9	243	15	US-10-195-895-366	Sequence 366, App	746	7	0.9	243	15	US-10-063-580-122	Sequence 122, App
674	7	0.9	243	15	US-10-199-302-366	Sequence 366, App	747	7	0.9	243	15	US-10-223-087-246	Sequence 246, App
675	7	0.9	243	15	US-10-201-323-366	Sequence 366, App	748	7	0.9	243	15	US-10-011-692A-352	Sequence 352, App
676	7	0.9	243	15	US-10-205-510-366	Sequence 366, App	749	7	0.9	243	15	US-10-011-692A-431	Sequence 431, App
677	7	0.9	243	15	US-10-205-891-366	Sequence 366, App	750	7	0.9	243	15	US-10-063-557-122	Sequence 122, App
678	7	0.9	243	15	US-10-206-917-366	Sequence 366, App	751	7	0.9	243	15	US-10-006-768A-352	Sequence 352, App
679	7	0.9	243	15	US-10-207-924-366	Sequence 366, App	752	7	0.9	243	15	US-10-006-768A-431	Sequence 431, App
680	7	0.9	243	15	US-10-207-924-366	Sequence 366, App	753	7	0.9	243	15	US-10-017-610A-352	Sequence 352, App
681	7	0.9	243	15	US-10-208-028-366	Sequence 366, App	754	7	0.9	243	15	US-10-017-610A-431	Sequence 431, App
682	7	0.9	243	15	US-10-063-538-122	Sequence 122, App	755	7	0.9	243	15	US-10-198-760-366	Sequence 366, App
683	7	0.9	243	15	US-10-012-121A-352	Sequence 352, App	756	7	0.9	243	15	US-10-201-772-366	Sequence 366, App
684	7	0.9	243	15	US-10-012-121A-431	Sequence 431, App	757	7	0.9	243	15	US-10-006-063A-352	Sequence 352, App
685	7	0.9	243	15	US-10-205-904-366	Sequence 366, App	758	7	0.9	243	15	US-10-006-063A-431	Sequence 431, App
686	7	0.9	243	15	US-10-175-753-366	Sequence 366, App	759	7	0.9	243	15	US-10-063-585-122	Sequence 122, App
687	7	0.9	243	15	US-10-180-553-366	Sequence 366, App	760	7	0.9	243	15	US-10-020-063A-352	Sequence 352, App
688	7	0.9	243	15	US-10-201-327-366	Sequence 366, App	761	7	0.9	243	15	US-10-020-063A-431	Sequence 431, App
689	7	0.9	243	15	US-10-121-062-366	Sequence 366, App	762	7	0.9	243	15	US-10-184-613-366	Sequence 366, App
690	7	0.9	243	15	US-10-063-599-122	Sequence 122, App	763	7	0.9	243	15	US-10-187-739-366	Sequence 366, App
691	7	0.9	243	15	US-10-006-116A-352	Sequence 352, App	764	7	0.9	243	15	US-10-206-907-366	Sequence 366, App
692	7	0.9	243	15	US-10-006-116A-431	Sequence 431, App	765	7	0.9	243	15	US-10-223-083-246	Sequence 246, App
693	7	0.9	243	15	US-10-006-117A-352	Sequence 352, App	766	7	0.9	243	15	US-10-015-391A-352	Sequence 352, App
694	7	0.9	243	15	US-10-006-117A-431	Sequence 431, App	767	7	0.9	243	15	US-10-015-391A-431	Sequence 431, App
695	7	0.9	243	15	US-10-017-527A-352	Sequence 352, App	768	7	0.9	243	15	US-10-183-009-366	Sequence 366, App
696	7	0.9	243	15	US-10-017-527A-431	Sequence 431, App	769	7	0.9	243	15	US-10-187-755-366	Sequence 366, App
697	7	0.9	243	15	US-10-183-003-366	Sequence 366, App	770	7	0.9	243	15	US-10-223-089-246	Sequence 246, App
698	7	0.9	243	15	US-10-183-016-366	Sequence 366, App	771	7	0.9	243	15	US-10-017-407A-352	Sequence 352, App
699	7	0.9	243	15	US-10-173-696-366	Sequence 366, App	772	7	0.9	243	15	US-10-017-407A-431	Sequence 431, App
700	7	0.9	243	15	US-10-013-913A-352	Sequence 352, App	773	7	0.9	243	16	US-10-006-041A-352	Sequence 352, App
701	7	0.9	243	15	US-10-013-913A-431	Sequence 431, App	774	7	0.9	243	16	US-10-006-041A-431	Sequence 431, App
702	7	0.9	243	15	US-10-063-595-122	Sequence 122, App	775	7	0.9	243	16	US-10-011-833A-352	Sequence 352, App
703	7	0.9	243	15	US-10-097-340-135	Sequence 135, App	776	7	0.9	243	16	US-10-011-833A-431	Sequence 431, App
704	7	0.9	243	15	US-10-125-923A-366	Sequence 366, App	777	7	0.9	243	16	US-10-015-822A-352	Sequence 352, App
705	7	0.9	243	15	US-10-176-491-366	Sequence 366, App	778	7	0.9	243	16	US-10-015-822A-431	Sequence 431, App
706	7	0.9	243	15	US-10-176-979-366	Sequence 366, App	779	7	0.9	243	16	US-10-063-588-122	Sequence 122, App
707	7	0.9	243	15	US-10-187-592-366	Sequence 366, App	780	7	0.9	245	14	US-10-045-992-2	Sequence 2, App1
708	7	0.9	243	15	US-10-007-194A-352	Sequence 352, App	781	7	0.9	247	11	US-09-029-047-4	Sequence 4, App1
709	7	0.9	243	15	US-10-007-194A-431	Sequence 431, App	782	7	0.9	256	15	US-10-156-761-13635	Sequence 13635, A
710	7	0.9	243	15	US-10-197-691-366	Sequence 366, App	783	7	0.9	261	12	US-09-849-138-47	Sequence 47, App1
711	7	0.9	243	15	US-10-198-771-366	Sequence 366, App	784	7	0.9	276	12	US-09-933-767-958	Sequence 958, App
712	7	0.9	243	15	US-10-013-430A-352	Sequence 352, App	785	7	0.9	276	15	US-10-023-282-958	Sequence 958, App
713	7	0.9	243	15	US-10-013-430A-431	Sequence 431, App	786	7	0.9	277	14	US-10-045-992-5	Sequence 5, App1
714	7	0.9	243	15	US-10-174-575A-366	Sequence 366, App	787	7	0.9	278	12	US-10-301-822-77	Sequence 77, App1
715	7	0.9	243	15	US-10-179-520-366	Sequence 366, App	788	7	0.9	278	12	US-10-124-805-515	Sequence 515, App
716	7	0.9	243	15	US-10-201-325-366	Sequence 366, App	789	7	0.9	278	14	US-10-007-805-515	Sequence 515, App
717	7	0.9	243	15	US-10-202-941-366	Sequence 366, App	790	7	0.9	278	15	US-10-076-622-515	Sequence 515, App
718	7	0.9	243	15	US-10-202-941-366	Sequence 366, App	791	7	0.9	278	15	US-10-060-036-4551	Sequence 4551, App
719	7	0.9	243	15	US-10-011-671A-352	Sequence 352, App	792	7	0.9	278	15	US-10-060-036-4560	Sequence 4560, App
720	7	0.9	243	15	US-10-011-671A-431	Sequence 431, App	793	7	0.9	278	15	US-10-097-340-131	Sequence 131, App
721	7	0.9	243	15	US-10-012-755A-352	Sequence 352, App	794	7	0.9	278	15	US-10-156-761-8795	Sequence 8795, App
722	7	0.9	243	15	US-10-012-755A-431	Sequence 431, App	795	7	0.9	278	15	US-10-177-293-496	Sequence 496, App
723	7	0.9	243	15	US-10-015-386A-352	Sequence 352, App	796	7	0.9	321	10	US-09-886-055-117	Sequence 117, App
724	7	0.9	243	15	US-10-015-386A-431	Sequence 431, App	797	7	0.9	321	11	US-09-804-291-117	Sequence 117, App
725	7	0.9	243	15	US-10-179-526-366	Sequence 366, App	798	7	0.9	322	12	US-10-017-161-884	Sequence 884, App
726	7	0.9	243	15	US-10-223-085-246	Sequence 246, App	799	7	0.9	322	15	US-10-156-761-10457	Sequence 10457, A
727	7	0.9	243	15	US-10-173-701-366	Sequence 366, App	800	7	0.9	415	12	US-10-238-075-1330	Sequence 1330, App
728	7	0.9	243	15	US-10-179-511-366	Sequence 366, App	801	7	0.9	427	12	US-09-813-432-8	Sequence 8, App1
729	7	0.9	243	15	US-10-179-518-366	Sequence 366, App	802	7	0.9	430	11	US-09-919-039-105	Sequence 105, App
730	7	0.9	243	15	US-10-183-018-366	Sequence 366, App	803	7	0.9	430	12	US-09-813-432-36	Sequence 36, App1
731	7	0.9	243	15	US-10-184-624-366	Sequence 366, App	804	7	0.9	430	14	US-10-026-001-1	Sequence 1, App1
732	7	0.9	243	15	US-10-184-657-366	Sequence 366, App	805	7	0.9	430	15	US-10-060-036-155	Sequence 155, App
733	7	0.9	243	15	US-10-197-701-366	Sequence 366, App	806	7	0.9	434	10	US-09-738-626-6682	Sequence 6682, App
734	7	0.9	243	15	US-10-197-706-366	Sequence 366, App	807	7	0.9	452	15	US-10-106-698-5102	Sequence 5102, App
735	7	0.9	243	15	US-10-201-857-366	Sequence 366, App	808	7	0.9	471	14	US-10-077-040-5	Sequence 5, App1
736	7	0.9	243	15	US-10-202-413-366	Sequence 366, App	809	7	0.9	471	15	US-10-160-293-5	Sequence 5, App1
737	7	0.9	243	15	US-10-202-938-366	Sequence 366, App	810	7	0.9	476	14	US-10-020-038-2	Sequence 2, App1
738	7	0.9	243	15	US-10-202-940-366	Sequence 366, App	811	7	0.9	504	15	US-10-156-761-14954	Sequence 14954, A
739	7	0.9	243	15	US-10-205-508-366	Sequence 366, App	812	7	0.9	525	11	US-09-882-694-8	Sequence 8, App1
740	7	0.9	243	15	US-10-205-905-366	Sequence 366, App	813	7	0.9	532	12	US-10-262-794A-30	Sequence 30, App1
741	7	0.9	243	15	US-10-206-918-366	Sequence 366, App	814	7	0.9	562	15	US-10-242-056-30	Sequence 30, App1
742	7	0.9	243	15	US-10-208-025-366	Sequence 366, App	815	7	0.9	595	12	US-09-849-138-2	Sequence 2, App1
743	7	0.9	243	15	US-10-223-084-246	Sequence 246, App	816	7	0.9	595	12	US-09-849-138-4	Sequence 4, App1
744	7	0.9	243	15	US-10-223-088-246	Sequence 246, App	817	7	0.9	595	12	US-09-849-138-6	Sequence 6, App1
745	7	0.9	243	15	US-10-223-090-246	Sequence 246, App	818	7	0.9	596	9	US-09-797-039-8	Sequence 8, App1

819	7	0.9	596	12	US-09-849-138-8	Sequence 8, Appli	892	6	0.8	73	11	US-09-991-262-46	Sequence 46, Appl
820	7	0.9	596	12	US-09-849-138-31	Sequence 31, Appl	893	6	0.8	73	12	US-10-321-856-293	Sequence 293, Appl
821	7	0.9	596	12	US-09-849-138-32	Sequence 32, Appl	894	6	0.8	73	15	US-10-142-835-4	Sequence 4, Appl
822	7	0.9	596	12	US-10-170-789-8	Sequence 8, Appli	895	6	0.8	74	9	US-09-764-853-712	Sequence 712, App
823	7	0.9	599	12	US-09-849-138-29	Sequence 29, Appl	896	6	0.8	74	9	US-09-764-898-233	Sequence 233, App
824	7	0.9	600	12	US-09-849-138-30	Sequence 30, Appl	897	6	0.8	74	11	US-09-764-881-108	Sequence 108, App
825	7	0.9	616	15	US-10-156-761-13507	Sequence 13507, A	898	6	0.8	74	15	US-10-073-865-71	Sequence 71, Appl
826	7	0.9	680	15	US-10-153-668-244	Sequence 244, App	899	6	0.8	75	15	US-10-106-698-6156	Sequence 6156, Ap
827	7	0.9	720	10	US-09-919-497-83	Sequence 83, Appl	900	6	0.8	76	14	US-10-016-157A-204	Sequence 204, Appl
828	7	0.9	852	10	US-09-738-626-6999	Sequence 6999, Ap	901	6	0.8	77	9	US-09-287-849-33	Sequence 33, Appl
829	7	0.9	974	15	US-10-156-761-9299	Sequence 9299, Ap	902	6	0.8	77	9	US-09-864-761-47447	Sequence 47447, A
830	7	0.9	1018	9	US-09-815-242-5186	Sequence 5186, Ap	903	6	0.8	77	12	US-10-359-460-33	Sequence 33, Appl
831	7	0.9	1070	12	US-10-174-677-102	Sequence 102, App	904	6	0.8	79	15	US-10-106-698-5623	Sequence 5623, Ap
832	7	0.9	1189	12	US-10-262-794A-26	Sequence 26, Appl	905	6	0.8	80	15	US-10-106-698-6303	Sequence 6303, Ap
833	7	0.9	1190	15	US-10-243-056-26	Sequence 26, Appl	906	6	0.8	80	15	US-10-156-761-7848	Sequence 7848, Ap
834	7	0.9	1311	8	US-08-954-701A-4	Sequence 4, Appli	907	6	0.8	81	10	US-09-738-626-6350	Sequence 6350, Ap
835	7	0.9	1311	12	US-09-754-032-4	Sequence 4, Appli	908	6	0.8	81	12	US-10-029-386-30350	Sequence 30350, A
836	7	0.9	1311	12	US-10-421-446-4	Sequence 4, Appli	909	6	0.8	81	14	US-10-086-623-17	Sequence 17, Appl
837	7	0.9	4303	11	US-09-904-968A-2	Sequence 2, Appli	910	6	0.8	81	15	US-10-260-539-17	Sequence 17, Appl
838	6	0.8	6	10	US-09-915-306-19	Sequence 19, Appl	911	6	0.8	85	10	US-09-796-692-692	Sequence 692, App
839	6	0.8	6	10	US-09-915-374-19	Sequence 19, Appl	912	6	0.8	85	15	US-10-040-862-692	Sequence 692, App
840	6	0.8	13	9	US-09-915-306-10	Sequence 10, Appl	913	6	0.8	87	11	US-09-774-639-305	Sequence 305, App
841	6	0.8	16	10	US-09-915-374-10	Sequence 10, Appl	914	6	0.8	87	11	US-09-969-730-232	Sequence 232, App
842	6	0.8	17	15	US-09-858-332-2	Sequence 2, Appli	915	6	0.8	90	11	US-09-764-872-316	Sequence 316, App
843	6	0.8	17	15	US-10-138-083-16	Sequence 16, Appl	916	6	0.8	91	11	US-09-843-676-220	Sequence 220, App
844	6	0.8	19	15	US-10-196-703-15	Sequence 15, Appl	917	6	0.8	91	11	US-09-838-486-220	Sequence 220, App
845	6	0.8	25	11	US-09-774-639-307	Sequence 307, App	918	6	0.8	91	15	US-10-053-758-220	Sequence 220, App
846	6	0.8	25	11	US-09-969-730-234	Sequence 234, App	919	6	0.8	91	15	US-10-054-295-220	Sequence 220, App
847	6	0.8	26	15	US-10-121-857-23	Sequence 23, Appl	920	6	0.8	91	15	US-10-054-611-220	Sequence 220, App
848	6	0.8	26	15	US-10-208-018-23	Sequence 23, Appl	921	6	0.8	91	15	US-10-156-761-11467	Sequence 11467, A
849	6	0.8	29	9	US-09-864-761-40009	Sequence 40009, A	922	6	0.8	94	15	US-10-227-629-11	Sequence 11, Appl
850	6	0.8	29	12	US-09-962-756-1357	Sequence 1357, Ap	923	6	0.8	94	15	US-10-156-761-14998	Sequence 14998, A
851	6	0.8	30	12	US-09-933-767-929	Sequence 929, App	924	6	0.8	98	12	US-10-065-133A-55	Sequence 55, Appl
852	6	0.8	30	15	US-10-083-357-1139	Sequence 1139, Ap	925	6	0.8	98	11	US-09-764-891-3671	Sequence 3671, Ap
853	6	0.8	30	15	US-10-023-282-929	Sequence 929, App	926	6	0.8	99	9	US-09-864-761-37490	Sequence 37490, A
854	6	0.8	37	9	US-09-864-761-46919	Sequence 46919, A	927	6	0.8	99	10	US-09-764-877-1608	Sequence 1608, Ap
855	6	0.8	38	12	US-10-315-964A-369	Sequence 369, App	928	6	0.8	99	15	US-10-106-698-4713	Sequence 4713, Ap
856	6	0.8	38	12	US-10-317-251A-369	Sequence 369, App	929	6	0.8	101	9	US-09-764-860-401	Sequence 401, App
857	6	0.8	39	12	US-10-317-252A-369	Sequence 369, App	930	6	0.8	101	15	US-10-074-095-401	Sequence 51, Appl
858	6	0.8	39	12	US-10-029-386-33330	Sequence 33330, A	931	6	0.8	105	10	US-09-795-006A-51	Sequence 59, Appl
859	6	0.8	41	9	US-09-864-761-47250	Sequence 47250, A	932	6	0.8	105	10	US-09-795-006A-59	Sequence 59, Appl
860	6	0.8	42	10	US-09-764-877-1588	Sequence 1588, Ap	933	6	0.8	105	10	US-09-795-006A-77	Sequence 77, Appl
861	6	0.8	43	9	US-09-864-761-44234	Sequence 44234, A	934	6	0.8	105	10	US-09-795-006A-153	Sequence 153, App
862	6	0.8	44	12	US-10-029-386-30666	Sequence 30666, A	935	6	0.8	105	10	US-09-795-006A-157	Sequence 157, App
863	6	0.8	45	14	US-10-011-445-49	Sequence 49, Appl	936	6	0.8	105	10	US-09-795-006A-159	Sequence 159, App
864	6	0.8	46	11	US-09-813-153-201	Sequence 201, App	937	6	0.8	105	10	US-09-795-006A-163	Sequence 163, App
865	6	0.8	47	9	US-09-864-761-40207	Sequence 40207, A	938	6	0.8	105	10	US-09-795-006A-165	Sequence 165, App
866	6	0.8	48	10	US-09-975-143-39	Sequence 39, Appl	939	6	0.8	105	10	US-09-795-006A-167	Sequence 167, App
867	6	0.8	48	10	US-09-975-143-39	Sequence 40, Appl	940	6	0.8	105	10	US-09-795-006A-169	Sequence 169, App
868	6	0.8	48	14	US-10-139-876-6	Sequence 6, Appli	941	6	0.8	105	10	US-09-795-006A-173	Sequence 173, App
869	6	0.8	49	9	US-09-864-761-44020	Sequence 44020, A	942	6	0.8	105	10	US-09-795-006A-175	Sequence 175, App
870	6	0.8	49	9	US-09-864-761-45127	Sequence 45127, A	943	6	0.8	105	14	US-10-001-876-187	Sequence 187, App
871	6	0.8	50	12	US-10-029-386-33298	Sequence 33298, A	944	6	0.8	106	11	US-09-764-891-4515	Sequence 4515, Ap
872	6	0.8	54	10	US-09-795-006A-81	Sequence 81, Appl	945	6	0.8	109	11	US-09-983-802-462	Sequence 462, App
873	6	0.8	55	10	US-09-925-300-1478	Sequence 1478, Ap	946	6	0.8	111	9	US-09-864-761-42542	Sequence 42542, A
874	6	0.8	55	15	US-10-034-934-123	Sequence 123, App	947	6	0.8	113	9	US-09-864-761-44575	Sequence 44575, A
875	6	0.8	56	11	US-09-229-173-42	Sequence 42, Appl	948	6	0.8	117	10	US-09-738-626-4398	Sequence 4398, Ap
876	6	0.8	57	12	US-10-238-075-443	Sequence 443, App	949	6	0.8	121	11	US-09-764-864-1125	Sequence 1125, Ap
877	6	0.8	59	8	US-08-424-550B-440	Sequence 440, App	950	6	0.8	121	11	US-09-975-719-421	Sequence 421, App
878	6	0.8	62	14	US-10-016-157A-151	Sequence 161, App	951	6	0.8	122	9	US-09-864-761-37388	Sequence 37388, A
879	6	0.8	63	12	US-10-029-386-29931	Sequence 29931, A	952	6	0.8	122	9	US-09-734-569-54	Sequence 54, Appl
880	6	0.8	64	11	US-09-764-891-3913	Sequence 3913, Ap	953	6	0.8	122	10	US-09-854-799-30	Sequence 30, Appl
881	6	0.8	64	15	US-10-145-415-36	Sequence 36, Appl	954	6	0.8	125	10	US-09-736-959A-21	Sequence 21, Appl
882	6	0.8	64	15	US-10-145-415-36	Sequence 38, Appl	955	6	0.8	125	10	US-09-854-799-54	Sequence 54, Appl
883	6	0.8	65	9	US-09-864-761-41837	Sequence 41837, A	956	6	0.8	126	10	US-09-795-006A-53	Sequence 53, Appl
884	6	0.8	66	9	US-09-864-761-47227	Sequence 47227, A	957	6	0.8	126	10	US-09-795-006A-55	Sequence 55, Appl
885	6	0.8	67	11	US-09-764-891-3101	Sequence 3101, Ap	958	6	0.8	126	10	US-09-795-006A-57	Sequence 57, Appl
886	6	0.8	67	12	US-10-029-386-33509	Sequence 33509, A	959	6	0.8	126	10	US-09-795-006A-61	Sequence 61, Appl
887	6	0.8	67	15	US-10-205-428-286	Sequence 286, App	960	6	0.8	126	10	US-09-795-006A-63	Sequence 63, Appl
888	6	0.8	71	9	US-09-925-299-1391	Sequence 1391, Ap	961	6	0.8	126	10	US-09-795-006A-65	Sequence 65, Appl
889	6	0.8	71	11	US-09-925-299-1391	Sequence 1391, Ap	962	6	0.8	127	10	US-09-795-006A-67	Sequence 67, Appl
890	6	0.8	73	9	US-09-216-393-293	Sequence 293, App	963	6	0.8	128	10	US-09-795-006A-69	Sequence 69, Appl
891	6	0.8	73	10	US-09-764-877-1872	Sequence 1872, Ap	964	6	0.8	128	10	US-09-795-006A-69	Sequence 69, Appl

965 6 0.8 128 10 US-09-795-006A-71 Sequence 71, Appl
966 6 0.8 128 10 US-09-795-006A-73 Sequence 73, Appl
967 6 0.8 128 10 US-09-795-006A-75 Sequence 75, Appl
968 6 0.8 128 10 US-09-795-006A-79 Sequence 79, Appl
969 6 0.8 129 10 US-09-795-006A-49 Sequence 49, Appl
970 6 0.8 130 12 US-10-325-717-7 Sequence 7, Appl
971 6 0.8 130 14 US-10-072-159-11 Sequence 11, Appl
972 6 0.8 130 14 US-10-072-159-12 Sequence 12, Appl
973 6 0.8 130 14 US-10-072-159-13 Sequence 13, Appl
974 6 0.8 130 14 US-10-060-845-11 Sequence 11, Appl
975 6 0.8 130 14 US-10-060-845-12 Sequence 12, Appl
976 6 0.8 130 14 US-10-060-845-13 Sequence 13, Appl
977 6 0.8 133 15 US-10-012-542-409 Sequence 409, App
978 6 0.8 133 16 US-10-286-421-31 Sequence 31, Appl
979 6 0.8 135 12 US-10-325-717-110 Sequence 110, App
980 6 0.8 135 12 US-10-325-717-111 Sequence 111, App
981 6 0.8 135 12 US-10-325-717-113 Sequence 113, App
982 6 0.8 136 10 US-09-938-497-8 Sequence 8, Appl
983 6 0.8 136 10 US-09-938-497-14 Sequence 14, Appl
984 6 0.8 136 12 US-10-342-224-106 Sequence 106, App
985 6 0.8 137 12 US-10-325-717-6 Sequence 30376, A
986 6 0.8 137 12 US-10-325-717-6 Sequence 6, Appl
987 6 0.8 138 11 US-09-847-208-155 Sequence 155, App
988 6 0.8 139 11 US-09-832-129-53 Sequence 53, Appl
989 6 0.8 139 11 US-09-272-975-51 Sequence 51, Appl
990 6 0.8 140 15 US-10-156-761-9549 Sequence 9549, App
991 6 0.8 142 15 US-10-227-616-103 Sequence 103, App
992 6 0.8 144 11 US-09-764-891-3838 Sequence 3838, App
993 6 0.8 144 12 US-10-325-717-5 Sequence 5, Appl
994 6 0.8 144 12 US-10-325-717-25 Sequence 25, Appl
995 6 0.8 144 12 US-10-325-717-41 Sequence 41, Appl
996 6 0.8 144 12 US-10-325-717-42 Sequence 42, Appl
997 6 0.8 144 12 US-10-325-717-43 Sequence 43, Appl
998 6 0.8 145 12 US-10-325-717-11 Sequence 27, Appl
999 6 0.8 145 12 US-10-325-717-27 Sequence 28, Appl
1000 6 0.8 145 12 US-10-325-717-28 Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-10-066-551-2
; Sequence 2, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045U51
; CURRENT APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR FILING DATE: 2002-01-31
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-08-06
; PRIOR FILING DATE: 2001-08-06
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-2

Query Match 23.1%; Score 175; DB 15; Length 764;
Best Local Similarity 100.0%; Pred. No. 2e-166;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPDGAHRTQARSFALSGNMYFTPOHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 532
Db 479 PLPDGAHRTQARSFALSGNMYFTPOHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 538
Qy 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFGNVIYAQTLDNDGRGPKSIEDDSEMKL 592
Db 539 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFGNVIYAQTLDNDGRGPKSIEDDSEMKL 598
Qy 593 VRYNQSADFYGABGEIYFXTPTPRYRIGVSGDYVRGRKLNLPSPGREDAYGNRP 647
Db 599 VRYNQSADFYGABGEIYFXTPTPRYRIGVSGDYVRGRKLNLPSPGREDAYGNRP 653

RESULT 2

US-10-204-887-121
; Sequence 121, Application US/10204887
; Publication No. US20030124569A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHANG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HULLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DAFFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1134 PCT
; CURRENT APPLICATION NUMBER: US/10/204,887
; CURRENT FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/185,216
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17;
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PERL Program
; SEQ ID NO 121
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:1072337.2.orf2:2000MAY01
US-10-204-887-121

Query Match 1.1%; Score 8; DB 15; Length 313;
Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 RVPAARLG 665
|||||
Db 55 RVPAARLG 62

RESULT 3

US-10-156-761-9479
; Sequence 9479, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9479
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9479

Query Match 1.1%; Score 8; DB 15; Length 387;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
|||||
Db 54 VSGELGLR 61

RESULT 4

US-09-996-008B-2
; Sequence 2, Application US/0996008B
; Publication No. US20030157677A1

GENERAL INFORMATION:

; APPLICANT: Mitsuhashi, Kazuya
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20030157677A1bhiro
; TITLE OF INVENTION: MUTANTS OF MYCOBACTERIUM VACCAR-DERIVED
; FILE REFERENCE: 14879-093001
; CURRENT APPLICATION NUMBER: US/09/996,008B
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: JP 2000-363894
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: JP 2001-254631
; PRIOR FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-996-008B-2

Query Match 1.1%; Score 8; DB 12; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179

Db 54 VSGELGLR 61
|||||

RESULT 5

US-10-286-326-4
; Sequence 4, Application US/10286326
; Publication No. US20030175903A1

GENERAL INFORMATION:

; APPLICANT: San, Ka-Yui
; APPLICANT: Berrios-Rivera, Susana
; APPLICANT: Bennett, George
; TITLE OF INVENTION: Recycling System for Manipulation of Intracellular NADH Availabi
; FILE REFERENCE: P02328US1
; CURRENT APPLICATION NUMBER: US/10/286,326
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/335,371
; PRIOR FILING DATE: 2001-11-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Pseudomonas
US-10-286-326-4

Query Match 1.1%; Score 8; DB 12; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
|||||
Db 54 VSGELGLR 61

RESULT 6

US-09-975-719-267
; Sequence 267, Application US/09975719
; Publication No. US20030022349A1

GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 267
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-267

Query Match 1.1%; Score 8; DB 11; Length 446;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 HRAVLGWR 245
|||||
Db 65 HRAVLGWR 72

RESULT 7

US-09-915-306-18
; Sequence 18, Application US/09915306
; Patent No. US20020081310A1

GENERAL INFORMATION:


```

; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-18

Query Match          0.9%; Score 7; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
DB 1 NLGDALD 7

```

```

RESULT 8
US-09-915-374-18
; Sequence 18, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-18

Query Match          0.9%; Score 7; DB 10; Length 7;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
DB 1 NLGDALD 7

```

```

RESULT 9
US-09-915-306-17
; Sequence 17, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-17

Query Match          0.9%; Score 7; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
DB 2 NLGDALD 8

```

```

RESULT 10
US-09-915-374-17
; Sequence 17, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-17

Query Match          0.9%; Score 7; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
DB 1 NLGDALD 7

```

QY 72 NLGDALD 78
Db 2 NLGDALD 8

RESULT 11

US-09-915-306-16
; Sequence 16, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-16

Query Match 0.9%; Score 7; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 3 NLGDALD 9

RESULT 12

US-09-915-374-16
; Sequence 16, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-16

Query Match 0.9%; Score 7; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 3 NLGDALD 9

RESULT 13

US-09-915-306-1
; Sequence 1, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-1

Query Match 0.9%; Score 7; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 4 NLGDALD 10

RESULT 14

US-09-915-374-1
; Sequence 1, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT


```
/ CURRENT FILING DATE: 2001-07-27
/ PRIOR APPLICATION NUMBER: 09/162,366
/ PRIOR FILING DATE: 1998-07-27
/ PRIOR APPLICATION NUMBER: SE 9703456-3
/ PRIOR FILING DATE: 1997-09-30
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 14
/ LENGTH: 11
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: epitope from
/ OTHER INFORMATION: mammalian protein
US-09-915-374-14

Query Match          0.9%; Score 7; DB 10; Length 11;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 5 NLGDALD 11

RESULT 19
US-09-915-306-12
/ Sequence 12, Application US/09915306
/ Patent No. US20020081310A1
/ GENERAL INFORMATION:
/ APPLICANT: Bjorklund, Viveka
/ APPLICANT: Bjorklund, Bertil
/ APPLICANT: Bjorklund, Peter
/ APPLICANT: Nap, Marius
/ APPLICANT: Ramaekers, Frans C.S.
/ APPLICANT: Schutte, Bert
/ TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
/ FILE REFERENCE: Bjorklund 09/162366
/ CURRENT APPLICATION NUMBER: US/09/915,306
/ CURRENT FILING DATE: 2001-07-27
/ PRIOR APPLICATION NUMBER: 09/162,366
/ PRIOR FILING DATE: 1998-09-29
/ PRIOR APPLICATION NUMBER: SE 9703456-3
/ PRIOR FILING DATE: 1997-09-30
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 12
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: epitope from
/ OTHER INFORMATION: mammalian protein
US-09-915-306-12

Query Match          0.9%; Score 7; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 6 NLGDALD 12

RESULT 20
US-09-915-374-12
/ Sequence 12, Application US/09915374
/ Publication No. US20020197657A1
/ GENERAL INFORMATION:
/ APPLICANT: Bjorklund, Viveka
/ APPLICANT: Bjorklund, Bertil
/ APPLICANT: Bjorklund, Peter
/ APPLICANT: Nap, Marius
```

```
/ APPLICANT: Ramaekers, Frans C.S.
/ APPLICANT: Schutte, Bert
/ TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
/ FILE REFERENCE: Bjorklund 09/162366
/ CURRENT APPLICATION NUMBER: US/09/915,374
/ CURRENT FILING DATE: 2001-07-27
/ PRIOR APPLICATION NUMBER: 09/162,366
/ PRIOR FILING DATE: 1998-07-27
/ PRIOR APPLICATION NUMBER: SE 9703456-3
/ PRIOR FILING DATE: 1997-09-30
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 12
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: epitope from
/ OTHER INFORMATION: mammalian protein
US-09-915-374-12

Query Match          0.9%; Score 7; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 6 NLGDALD 12

RESULT 21
US-09-915-306-13
/ Sequence 13, Application US/09915306
/ Patent No. US20020081310A1
/ GENERAL INFORMATION:
/ APPLICANT: Bjorklund, Viveka
/ APPLICANT: Bjorklund, Bertil
/ APPLICANT: Bjorklund, Peter
/ APPLICANT: Nap, Marius
/ APPLICANT: Ramaekers, Frans C.S.
/ APPLICANT: Schutte, Bert
/ TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
/ FILE REFERENCE: Bjorklund 09/162366
/ CURRENT APPLICATION NUMBER: US/09/915,306
/ CURRENT FILING DATE: 2001-07-27
/ PRIOR APPLICATION NUMBER: 09/162,366
/ PRIOR FILING DATE: 1998-09-29
/ PRIOR APPLICATION NUMBER: SE 9703456-3
/ PRIOR FILING DATE: 1997-09-30
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 13
/ LENGTH: 13
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: epitope from
/ OTHER INFORMATION: mammalian protein
US-09-915-306-13

Query Match          0.9%; Score 7; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 7 NLGDALD 13

RESULT 22
US-09-915-374-13
/ Sequence 13, Application US/09915374
/ Publication No. US20020197657A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-13

Query Match          0.9%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 7 NLGDALD 13

RESULT 23
US-09-915-306-11
; Sequence 11, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-11

Query Match          0.9%; Score 7; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14

US-09-915-306-3
; Sequence 3, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-3

Query Match          0.9%; Score 7; DB 9; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14
```

```

RESULT 24
US-09-915-374-11
; Sequence 11, Application US/09915374
; Publication NO. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-11

Query Match          0.9%; Score 7; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14

RESULT 25
US-09-915-306-3
; Sequence 3, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-3

Query Match          0.9%; Score 7; DB 9; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14
```

```

US-09-915-306-2
Query Match      0.9%; Score 7; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      7 NLGDALD 13

RESULT 26
US-09-915-374-3
; Sequence 3, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-3

Query Match      0.9%; Score 7; DB 10; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      8 NLGDALD 14

RESULT 27
US-09-915-306-2
; Sequence 2, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-2
Query Match      0.9%; Score 7; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      7 NLGDALD 13

RESULT 28
US-09-915-374-2
; Sequence 2, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-2

Query Match      0.9%; Score 7; DB 12; Length 46;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      7 NLGDALD 13

RESULT 29
US-10-124-805-518
; Sequence 518, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 518
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-124-805-518

Query Match      0.9%; Score 7; DB 12; Length 46;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 AGLVDVA 159
 Db 9 AGLVDVA 15

RESULT 30

US-10-007-805-518
 ; Sequence 518, Application US/10007805
 ; Publication No. US20020150581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Durham, Margarita
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.470C10
 ; CURRENT APPLICATION NUMBER: US/10/007.805
 ; CURRENT FILING DATE: 2001-12-07
 ; NUMBER OF SEQ ID NOS: 593
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 518
 ; LENGTH: 46
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-007-805-518

Query Match 0.9%; Score 7; DB 14; Length 46;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 AGLVDVA 159
 Db 9 AGLVDVA 15

RESULT 31

US-10-076-622-518
 ; Sequence 518, Application US/10076622
 ; Publication No. US20030023036A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Sleath, Paul R.
 ; APPLICANT: Persing, David H.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.470C11
 ; CURRENT APPLICATION NUMBER: US/10/076.622
 ; CURRENT FILING DATE: 2002-02-13
 ; NUMBER OF SEQ ID NOS: 627
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 518
 ; LENGTH: 46
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-076-622-518

Query Match 0.9%; Score 7; DB 15; Length 46;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 AGLVDVA 159
 Db 9 AGLVDVA 15

RESULT 32

US-09-864-761-42772
 ; Sequence 42772, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 42772
 ; LENGTH: 47
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AL096678.8
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.89
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
 US-09-864-761-42772

Query Match 0.9%; Score 7; DB 9; Length 47;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 33

US-09-933-767-963

Sequence 963, Application US/09933767

Publication No. US20030181692A1

GENERAL INFORMATION:

APPLICANT: Ni et al.

TITLE OF INVENTION: 207 Human Secreted Proteins

FILE REFERENCE: P2007P2

CURRENT APPLICATION NUMBER: US/09/933.767

CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: PCT/US01/05614

PRIOR FILING DATE: 2001-02-21

PRIOR APPLICATION NUMBER: 60/184,836

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/193,170

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: 09/205,258

PRIOR FILING DATE: 1998-12-04

PRIOR APPLICATION NUMBER: PCT/US98/11422

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/048,885

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/049,375

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,881

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,880

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,896

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/049,020

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,876

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,895

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,884

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,894

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,971

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,964

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,882

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,899

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,893

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,900

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,901

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,892

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,915

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/049,019

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,970

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,972

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/048,916

PRIOR FILING DATE: 1997-06-06

PRIOR APPLICATION NUMBER: 60/049,373

PRIOR FILING DATE: 1997-06-06

```

; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT FILING DATE: 2001-12-20
; EARLIER APPLICATION NUMBER: US/10/023,282
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 09/205,258
; EARLIER FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898

```

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 963
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-282-963

Query Match          0.9%; Score 7; DB 15; Length 51;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
DB 14 AGLVDVA 20
|||||

RESULT 35
US-10-156-761-15079
; Sequence 15079, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15079
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15079

Query Match          0.9%; Score 7; DB 15; Length 52;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 TLLYSSG 150
DB 7 TLLYSSG 13
|||||

RESULT 36
US-10-156-761-7973
; Sequence 7973, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```

```

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7973
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7973

Query Match      0.9%; Score 7; DB 15; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      446 GGVREK 452
DB      28 GGVREK 34

RESULT 37
US-09-764-868-801
; Sequence 801, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND ANTIBODIES
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 801
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-801

Query Match      0.9%; Score 7; DB 10; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      489 LSGNWF 495
DB      26 LSGNWF 32

RESULT 38
US-10-123-058-6
; Sequence 6, Application US/10123058
; Publication No. US20030194695A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: MCGUIRE, MICHAEL J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES USEFUL AS
; TITLE OF INVENTION: ADJUVANTS
; FILE REFERENCE: UTSD:664US
; CURRENT APPLICATION NUMBER: US/10/123,058
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 6
; LENGTH: 136
; TYPE: PRT
; ORGANISM: PARAPOX
US-10-123-058-6

Query Match      0.9%; Score 7; DB 12; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      42 SRPRATS 48
DB      112 SRPRATS 118

RESULT 39
US-10-414-609-6
; Sequence 6, Application US/10414609
; Publication No. US20030194737A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: MCGUIRE, MICHAEL J.
; TITLE OF INVENTION: Use of Parapox B2L Protein to Modify Immune Responses to Adminis
; TITLE OF INVENTION: Antigens
; FILE REFERENCE: UTSD:664-2US
; CURRENT APPLICATION NUMBER: US/10/414,609
; CURRENT FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: US 10/123,058
; PRIOR FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: US 60/336,694
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 136
; TYPE: PRT
; ORGANISM: PARAPOX
US-10-414-609-6

Query Match      0.9%; Score 7; DB 12; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      42 SRPRATS 48
DB      112 SRPRATS 118

RESULT 40
US-10-153-668-364
; Sequence 364, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, GOICHI
; APPLICANT: MATSUDA, AKIO
; APPLICANT: MURAMATSU, SHUJI
; APPLICANT: ISHIZAWA, KENYA
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10

```

```
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 364
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-364
```

```
Query Match          0.9%; Score 7; DB 15; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 FTPQKLL 501
Db 92 FTPQKLL 98
```

```
Search completed: November 14, 2003, 11:13:40
Job time : 37 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 16:20:31 ; Search time 4856 Seconds
(without alignments)
6385.811 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 4036
Sequence: 1 MAQTILKPIVLSILLINTPL.....FLSDTQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 288711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO_spool/US9936377/runat_14112003_104449_22108/app_query.fasta_1.903
-DB=GenEmbl -OPMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US9936377_@CGN_1_1_3078_@runat_14112003_104449_22108 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.ov.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4036	100.0	2277	6	AX035174	Sequence
2	3928	97.3	2295	6	AX548017	Sequence
C 3	3928	97.3	331801	1	NMA422491	Neisseria
4	3927	97.3	2277	6	AX300375	Sequence
C 5	3927	97.3	10700	1	AE002448	Neisseria
C 6	3927	97.3	349980	6	AX044031	Sequence
7	1725.5	42.8	4931	1	AY028475	Sequence
8	1585	39.3	2457	6	AX052567	Sequence
9	1585	39.3	2646	6	AX394914	Sequence
10	1585	39.3	89047	6	AX067459	Sequence
C 11	1492.5	37.0	10120	1	AE006149	Sequence
C 12	1364	33.8	11375	1	AE006179	Sequence
C 13	1307	32.4	10166	1	AE006112	Sequence
14	1205.5	29.9	2942	6	AX394912	Sequence
15	1205.5	29.9	269223	6	AX067466	Sequence
16	1202	29.8	2742	6	AX052569	Sequence
17	745	18.5	10947	1	AE012378	Xanthomon
18	718.5	17.8	11947	1	AE011924	Xanthomon
19	694.5	17.2	10158	1	AE003890	Xanthomon
C 20	691.5	17.1	295250	1	AX321862	Nitrosomo
C 21	689.5	17.1	300885	1	AE012559	Xylella f
C 22	679	16.8	15470	1	AE004513	Pseudomon
C 23	642	15.9	301995	1	AE016779	Pseudomon
C 24	636	15.8	11626	1	AE005695	Caulobact
C 25	469	11.6	300400	1	AP005943	Bradyrhiz
C 26	396	9.8	10214	1	AE011477	Leptospir
C 27	392	9.7	303750	1	AE016931	Bacteroid
28	354.5	8.8	11549	1	AE006167	Pasteurel
29	290	7.2	17725	1	AE004349	Vibrio ch
30	271.5	6.7	1173	6	AX079015	Sequence
31	265	6.6	10161	1	AE005425	Escherich
32	265	6.6	286857	1	AP002559	Escherich
33	265	6.6	311143	1	AE016762	Escherich
C 34	264.5	6.6	2589	1	AF337562	Escherich
C 35	255	6.3	16841	1	AE004749	Pseudomon
C 36	253	6.3	10777	1	AE006093	Pasteurel
C 37	244.5	6.1	10399	1	AY008342	Plesiomon
38	244.5	6.1	12558	1	AB015838	Shewanell
C 39	243	6.0	16256	1	AE000694	Aquifex a
40	240.5	6.0	336182	1	AE016813	Vibrio vu
41	239	5.9	2501	1	AF127222	Pseudomon
42	238.5	5.9	7702	1	AF055999	Pseudomon
43	236.5	5.9	13223	1	AE004885	Pseudomon
44	236.5	5.9	298050	1	EX321861	Nitrosomo
C 45	236	5.8	335862	1	AP005089	Vibrio pa

ALIGNMENTS

RESULT 1


```

Qy 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
Db 1681 CTCTACGGCAACCGCTCGGCAACTACATTTAGCCCAACCTTAAACGACGACGCGGC 1740
Qy 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAla 600
Db 1741 CCCAAATCCATCGAAGACGACGAGAAATGAGCTCGTGCGCTACACCAATCCGCTGCG 1800
Qy 601 AspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgGlyGly 620
Db 1801 GACTTCTACGCGCGGAGGCGGAATCTACTTCAACCGACACCGCTACCGCATCGGC 1860
Qy 621 ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGlu 640
Db 1861 GTTTCGCGCGACTATGTACGAGGCGCTGTGAAACCTGCGCTACCGCGGAGGAA 1920
Qy 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
Db 1921 GATGCTACGGACCGACCTCTTCTATCGCGGACGACGACCAACCGCCCTCGCGTTCCG 1980
Qy 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
Db 1981 GCTGCGCGCTCGGCTTCCACTGAAAGCTCGCTACCGACCGCATCGATGCCAATTG 2040
Qy 681 AspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db 2041 GACTACTACCGGTGTTTGCCCAACAACTCGCCCGCTACGAAACGCGCACGCCCGGA 2100
Qy 701 HisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsn 720
Db 2101 CACCATATGCTCAACTCGCGCGCAACTACCGCGCAATACGCTATGGCGAGTGAAT 2160
Qy 721 TrpTyrValLysAlaAspAsnLeuLysGlnSerValTyrAlaHisSerSerPheLeu 740
Db 2161 TGGTACGTCAAGCGCAGCAACCTGCTCAACCAATCCGTTTACGCGCCACAGCAGCTTCTC 2220
Qy 741 SerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2221 TCTGATACGCCCAATGGCGCAGCTTTACCGGTGCGGTAAACGTGAAGTTT 2274

RESULT 2
AX548017 2295 bp DNA linear PAT 26-NOV-2002
LOCUS Sequence 7 from Patent WO02060936.
DEFINITION AX548017
ACCESSION AX548017
VERSION AX548017.1 GI:25813126
KEYWORDS
SOURCE
ORGANISM
Neisseria gonorrhoeae
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 Apicella, M.A., Edwards, J.L., Gibson, B.W., Scheffler, K. and Brown, E.
AUTHORS Vaccine and compositions for the prevention and treatment of
TITLE neisserial infections
JOURNAL UNIVERSITY OF IOWA RESEARCH FOUNDATION (US) ; University of
Patent: WO 02060936-A 7 08-AUG-2002;
California, Los Angeles (US) ; Apicella, Michael A. (US) ; Edwards,
Jennifer, L. (US) ; Gibson, Bradford W. (US) ; Scheffler, Karoline
(US)

FEATURES
Location/Qualifiers
1..2295
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/db_xref="taxon:485"
BASE COUNT 606 a 553 g 412 t
ORIGIN

Alignment Scores:
Pred. No.: 6.4e-229 Length: 2295
Score: 3928.00 Matches: 744
Percent Similarity: 98.68% Conservative: 5
Best Local Similarity: 98.02% Mismatches: 9

```

```

Query Match: 97.32% Indels: 2
DB: 6 Gaps: 0
US-09-936-377-2 (1-758) x AX548017 (1-2295)
Qy 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuThrProLeu 20
Db 19 ATGGCAACAACCTACACTCAACCCCATTTGTTTATCAATTCCTTTAATCAACACACCCCTC 78
Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
Db 79 CTCCTCCACAGCGCATGGAACCTGACCAATCAGTGGGCTTGAAACCGGTGACGCTCGTGGC 138
Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 139 AAAAGCGCTCGCGGCCACTTCGGGGCTGCTGCACACTTCTACCGCTTCCGACAAAATC 198
Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 199 ATCAGCGCGGACACCTTGGCAAAAAGCCGCACTTGGGTGATGCTTTAGACGGCGTA 258
Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db 259 CCGGGCATTCATGCTCGCAATAGCGCGCGCGCATCCGCTCCGCTTATTCGGGTCAA 318
Qy 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 319 ACAGGCGACGCGAATTAAAGTGTGAACCATCAACGCGGAAACGGCGACATGGCGGACTTC 378
Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
Db 379 TCTCCAGACCATGCAATCATGGTGGACAGCGGCTTGTCCCAACAGCTCGAATCCTGGCG 438
Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 439 GGTCCGGTTACGCTCTTGTACAGCTCGGCAATGTGCGGGGTGCTGCTCATGTTGCGCAT 498
Qy 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 180
Db 499 GGCATAATCCCGAAAAATGCTGAAACGGGCTATCGGCGCACTCGATTCGATTCGTTTG 558
Qy 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db 559 AGCAGCGGCAATCTGGAATAACTCACTCGCGCGCATCAATATCGTTTGGSCAAAAC 618
Qy 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 619 TTTGTATTGCACACGGAAGGCTGTACCGCAATCGGGGATTCACCGCTACCGCGTTAC 678
Qy 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 679 CGCAATCTGAACCGCTCGCGGACAG-CCACGCGCATTCGCAACCGGCGACATCGGCT 737
Qy 241 ValLeuGlyTrpArgLysArgPheTyrArgArgThr-TyrSerAspArgAspGlnTy 260
Db 738 GTCTTGGGTTGGGAAAAAGGCTTTATCGCGCGCATATCAGCGACCGCTCGCGCAATA 797
Qy 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLy 280
Db 798 TGTCTGCTGCCCCACGACCCAGCAATGATGATTCGACCGCCGACATCATCTGGCAAA 857
Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
Db 858 GAGTTTGTATTACAAACGCTATTTCAGCTTTTATTCGCACTTTTACCGACGACGACAT 917
Qy 300 LAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
Db 918 CGATTACGCAATCCCGGCTTGGCTGCGGCTTTTCAGCGACGATGATGACACGCCCA 977
Qy 320 sAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGl 340
Db 978 TGCCCAACAGCGCAAAACCTTGGATGACCTCGCGCAACAAACGCTACGAACTCCGCGCGA 1037

```


QY 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
 |||||
 Db 1038 ATGGAAGCAACCGTTCCCGGTTTGAAGCCCTGCGGTACACCTGGAACCGCAACGACTA 1097
 |||||
 QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 |||||
 Db 1098 CGCCACGACGAAAGCAGCGAGTGCAGTAGAAAACCTTTTAAACACCAACGCAAAA 1157
 |||||
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGI 400
 |||||
 Db 1158 CGCCCGTATCGAGTTGCGGCCACCAACCCATAGCCGCTCTGAAAGGCGAGCTGGGGGTGCA 1217
 |||||
 QY 400 nTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
 |||||
 Db 1218 ATATTGGGACAAAATCCAGTGTCTTATCTGCCCATCCGAGCGGTCAACCAACCGAT 1277
 |||||
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 |||||
 Db 1278 GCTGCTTGACATTAAGTGCACATTTACAGCTTTTTCGGTGTAGAACAGGCAACCTGGGA 1337
 |||||
 QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 |||||
 Db 1338 CAACTTCACGCTTGAAGCGGGGTAGCCGTGGAACAAACAAAGGCTCCATCCGTACGA 1397
 |||||
 QY 460 pLysAlaLeuIleAspArgGluAsnTyrTrpLysGlnProLysProLeuProAspLeuGlyAlaHi 480
 |||||
 Db 1398 CAAAGCATTCATTGATCGGGAACAACTACTACAACCATCCCTGCCCGCAGACTCGGCGCGCA 1457
 |||||
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTrpPheThrProGlnHisIy 500
 |||||
 Db 1458 CGCCCAACCGCCGCTCAATTCGCACTTTCCGGCAACTGGTATTTACGCCCAACACACAA 1517
 |||||
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 |||||
 Db 1518 ACTCAGCTGACCGCTCCCATCAGGAACGCTCGCTCAACGCAAGAGCTGTACGCACA 1577
 |||||
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 |||||
 Db 1578 CGCAACACAGCTCGCCACCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG 1637
 |||||
 QY 540 sSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAl 560
 |||||
 Db 1638 TTCCAACATATCGAATCGCGCTGGGCTACGAGGCGGACCGCTGGCAATACAACTGGC 1697
 |||||
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGI 580
 |||||
 Db 1698 ACTCTACCGCAACCGCTTCGGCAACTACATTTACGCCCAACACCTTAAACGACGACGCGG 1757
 |||||
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 |||||
 Db 1758 CCCCAAATCATCGAAGACGACAGCGAATGAAGCTCGTGGCTACAAACCAATCCGGTGC 1817
 |||||
 QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGI 620
 |||||
 Db 1818 GGACTTCTACGGCGGAGCGGAAATCTACTTTCAACCGCACACCGCGCTACCGCATCGG 1877
 |||||
 QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGI 640
 |||||
 Db 1878 CGTTTCGGCGCATATGTACAGCGCTCTGAAAAACCTTCCCTTCCCTACCGCGCAGGA 1937
 |||||
 QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 |||||
 Db 1938 AGACGCTACGGCAACCGCCCACTCATTTGCCAAGCGCACAAACCGCCCTCGCGTTCC 1997
 |||||
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnIle 680
 |||||
 Db 1998 GCGTCGCGGCTCGCGCTCCACTTGAAAGCTCGCTGACCGACCGCATGATGCCAATTT 2057
 |||||
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGI 700
 |||||
 Db 2058 GGACTACTACGGGTGTTCGCCCAACAAACCTCGCCCGCTACGAAACGCGACGCGCCGG 2117
 |||||
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAs 720
 |||||

Db 2118 ACACCATATGCTCAACTCGCGCAAACTACCGCGCAATACCGCTATATGCGAGTGGAA 2177
 |||||
 QY 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheIe 740
 |||||
 Db 2178 TTGGTACGTCAAAAGCGGACAAACCTGTCTCAACCAATCGTTTACGCCACACGACGACTTCT 2237
 |||||
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 |||||
 Db 2238 CTCTGATACCGCGCAATATGGCGCGAGCTTTACCGCGCGCTGAACGTGAAGTTT 2292
 |||||
 RESULT 3
 NMA422491/c 331801 bp DNA linear BCT 02-SEP-2002
 LOCUS Neisseria meningitidis serogroup A strain Z2491 complete genome;
 DEFINITION segment 4/7.
 VERSION AL162755 AL157959
 KEYWORDS AL162755.2 GI:7379742
 SOURCE Neisseria meningitidis Z2491
 ORGANISM Neisseria meningitidis Z2491
 Bacteria: Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 1 (bases 1 to 331801)
 REFERENCE Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
 AUTHORS Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
 Davies,S.R., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
 Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
 Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
 Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
 Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE 2 (bases 1 to 331801)
 AUTHORS Parkhill,J.
 DIRECT SUBMISSION
 TITLE Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
 JOURNAL sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
 NOTES:
 Details of N. meningitidis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
 FEATURES
 source
 1. 331801
 /organism="Neisseria meningitidis Z2491"
 /mol_type="genomic DNA"
 /strain="Z2491"
 /db_xref="taxon:122587"
 /note="serogroup: A"
 117. .1333
 /gene="NMA1040"
 /pseudo
 117. .1333
 /gene="NMA1040"
 /note="NMA1040"
 /note="possible type I restriction-modification
 system specificity protein, pseudogene, len: 1217 bp;
 N-terminus shows weak similarity to the C-terminal half of
 TR:Q50359 (EMBL:L25415), hsdS18, Mycoplasma pulmonis
 restriction-modification enzyme subunit S1B (336 aa),
 fasta scores; E(): 2.3e-06, 26.2% identity in 141 aa
 overlap and to many hypothetical restriction-modification
 subunits. Also similar to NMA1041, fasta scores; E():
 4.8e-10, 37.0% identity in 100 aa overlap. C-terminus
 similar to part of SW:fls1 ECOLI (EMBL:X13145), hsdS,
 Escherichia coli type I restriction enzyme EcoR124II
 specificity protein (410 aa), fasta scores; E(): 1.9e-14,
 39.6% identity in 182 aa overlap. Similar to NMA1040, E():
 2.9e-09, 38.0% identity in 100 aa overlap. Contains pfam
 match to entry PF01420 Methylase_S, Type I restriction
 modification DNA specificity domain. Contains a G(8) tract

which would allow translation as an intact CDS, if variable. Lies within a region of unusually low GC content"

```
/pseudo
/codon_start=1
/transl_table=11
/product="pseudogene (putative type I
restriction-modification system specificity protein)"
complement(157. .166)
```

misc_feature

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

misc_feature

```
/label=DUS
```

```
/gene="NMA1040"
```

```
/note="Pfam match to entry PF01420 Methylase_S, Type I
restriction modification DNA specificity domain, score
105.50, E-value 1.1e-27"
```

repeat_region

```
/pseudo
699. .706
```

misc_feature

```
/note="(g)8"
```

```
726. .735
```

```
/gene="NMA1040"
```

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/pseudo
```

misc_feature

```
/label=DUS
```

```
740. .1231
```

```
/gene="NMA1040"
```

```
/note="Pfam match to entry PF01420 Methylase_S, Type I
restriction modification DNA specificity domain, score
158.70, E-value 1e-43"
```

```
/pseudo
```

```
1405. .1408
```

```
1415. .4509
```

```
/gene="NMA1042"
```

```
/pseudo
```

```
1415. .4509
```

```
/gene="NMA1042"
note="NMA1042, pseudogene, probable type I
restriction-modification system restriction protein, len:
291 aa; similar to parts of many e.g. SW:TR1L_ECOLI
(EMBL:X13145), hsdR, Escherichia coli type I restriction
enzyme EcoRI2411 R protein (EC 3.1.21.3) (1033 aa), fasta
scores; E(): 0, 73.9% identity in 291 aa overlap (frame
1), followed by a stop codon, E(): 0, 92.9% identity in
127 aa overlap (frame 1), followed by a frameshift near a
poly-A tract, E(): 0, 72.5% identity in 604 aa overlap
(frame 3)"
/pseudo
```

```
/codon_start=1
```

```
/transl_table=11
```

```
/product="type I restriction-modification system
restriction protein (pseudogene)"
```

```
complement(1642. .1651)
```

misc_feature

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/label=DUS
```

```
1687. .1696
```

misc_feature

```
/gene="NMA1042"
```

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/pseudo
```

```
1761. .1770
```

misc_feature

```
/gene="NMA1042"
```

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/pseudo
```

```
3906. .3915
```

misc_feature

```
/gene="NMA1042"
```

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/pseudo
```

```
/label=DUS
```

stem_loop

```
4514. .4545
```

```
/note="stem loop containing DNA uptake sequences: aaaat
gccgtctgaa at ttacagcggc atttt"
```

misc_feature

```
4519. .4528
```

misc_feature

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/label=DUS
```

```
complement(4531. .4540)
```

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/label=DUS
```

```
complement(4559. .6838)
```

```
/gene="clpA"
```

```
complement(4559. .6838)
```

```
/gene="clpA"
```

```
complement(4559. .6838)
```

```
/note="NMA1045, clpA, probable ATP-dependent protease
ATP-binding protein, len: 759 aa; similar to many e.g.
SW:CLPA_ECOLI (EMBL:M31045), clpA, Escherichia coli
```

```
ATP-dependent Clp protease ATP-binding subunit (758 aa),
fasta scores; E(): 0, 56.2% identity in 762 aa overlap.
```

```
Similar to NMA1683, fasta scores; E(): 0, 37.2% identity
```

```
in 844 aa overlap. Contains Pfam match to entry PF00495
```

```
clpA_B, Chaperonin clpA/B, PS00870 Chaperonins clpA/B
```

```
signature 1, PS00871 Chaperonins clpA/B signature 2 and
```

```
two PS00017 ATP/GTP-binding site motif A (P-loop)"
```

```
/codon_start=1
```

```
/transl_table=11
```

```
/product="putative ATP-dependent protease ATP-binding
```

```
protein"
```

```
/protein_id="CAB84311.1"
```

```
/db_xref="GI:7379743"
```

```
/db_xref="SPTREMBL:Q9JV15"
```

```
/translation="MLSPEHLIQLLYREARNARYBFISLEHLLVLIBEAAVNVN
```

```
LKICGDLKVLSEOLAAVAENTPOIPDHLLDTVETQTFQFVORAMVHTQSAGK
```

```
AAEPLDILVAMSESSESHAVYPLKLOSLTRFEVLRCIAHGSPDEGNDGDLGREG
```

```
EEAQKTSLSLDYTVNLNAEKAGRIDPLIGRKHEMERLVQILCRRRKNLLVGEAG
```

```
VGTALAEGLAHQIVKGDIEDALDKADVAIDMGSLLAGIKYRGDFEARKSVLKQLE
```

```
KIPHAIFIIDEIHTIIGAGSTGMDANLLKALAKALRCICGATTYDYRTIFDK
```

```
DHALSRFQKIDVVEPTVSETVQILRGKPMFEGFHVRYTQGALEAAELSAARYNE
```

```
RFLPDKAIDVDEAGAAQRIPLPKSKQKKVIGKQIETVIAKVARIPEKTVSHDKQVL
```

```
QFLGRDLKNMVGQENAIYALVAAMKRSGLGLPDKPIGSLFSGPTGVGKTEVAKO
```

```
LAVSMGVLPQRFQMSYMERHARSLIGAPGYGVFGQGLLTFEANKQPKCVLLLDE
```

```
IEKHPDIFNVLLQVMDAGKLTNNKSGADFNNVILMTNAGAESLRSGISGTFKR
```

```
ERGDMAINKLFTPEFNRDLAIPFAPLSEPIIVKVDKFLQLHEHLLDKKVEAE
```

```
FTPALRYLAEKGFQPMQGRPMNRLLIQEKIRKPLADELLFGKLVDGSGFVRIDWAAK
```

```
EEAVLPKKSQKVPETETV"
```

```
complement(4637. .5578)
```

```
/gene="clpA"
```

```
/note="Pfam match to entry PF00495 clpA_B, Chaperonin
```

```
clpA/B, score 583.70, E-value 1.1e-171"
```

```
complement(5228. .5284)
```

```
/gene="clpA"
```

```
/note="PS00871 Chaperonins clpA/B signature 2"
```

```
complement(5330. .5353)
```

```
/gene="clpA"
```

```
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
```

```
complement(5894. .5932)
```

```
/gene="clpA"
```

```
/note="PS00870 Chaperonins clpA/B signature 1"
```

```
complement(6173. .6196)
```

```
/gene="clpA"
```

```
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
```

```
6374. .6383
```

```
/note="Core DNA uptake sequence: gccgtctgaa"
```

```
/label=DUS
```

```
complement(6842. .7153)
```

```
/gene="NMA1046"
```

```
complement(6842. .7153)
```

```
/gene="NMA1046"
```

```
/note="NMA1046, len: 103 aa; unknown, similar to bacterial
```

```
hypothetical proteins e.g. SW:YLJA_ECOLI (EMBL:AE000190),
```

```
YljA, Escherichia coli hypothetical protein (103 aa),
```

```
fasta scores; E(): 7.5e-14, 44.6% identity in 83 aa
```

Alignment Scores:
Pred. No.: 3,53e-226 Length: 331901
Score: 3928.00 Matches: 744
Percent Similarity: 98.68% Conservative: 5
Best Local Similarity: 98.02% Mismatches: 9

Query Match:	97.32%	Indels:	2
DB:	1	Gaps:	0
US-09-936-377-2 (1-758) x NMA422491 (1-331801)			
QY	1	MetAlaGlnThrThrLeuLysProlLeuValLeuSerIleLeuLeuIleAsnThrProLeu	20
DB	115608	ATGGACAACTACACTCAACCCCATGTTTATCAATTCATCAACACACCCCTC	115549
QY	21	LeuAlaGlnAlaHisGlnThrGluGlnSerValGlyLeuGluThrValThrValValGly	40
DB	115548	CTCTCCCAAGCGCATGGAACTGAGCAATCAGTGGCTTGGAAACGGTCAGCGTCGCGC	115489
QY	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
DB	115488	AAAAGCGTCGCGGCGCACTTCGGGGCTGTCGACACTTCCTACCGCCTCCGACAAATC	115429
QY	61	IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal	80
DB	115428	ATCAGCGCGCACCTTCGCAAAAAAGCCGTCACCTTGGGTGATGCTTTAGACGGCGTA	115369
QY	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln	100
DB	115368	CGGGCAATTCATGCTCGCAATACGGCGCGCGCATCCGCTCCCGTTATTTCGGGTCAA	115309
QY	101	ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe	120
DB	115308	ACAGGCAGACGGATTAAAGTGTGAACCATCAACGGCGAAACGGCGGACATGGCGGACTTC	115249
QY	121	SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg	140
DB	115248	TCTCAGACCATCGAATCATGTGTGACAGCGCTTGTGCCACAGTGCAGATCTCGCGC	115189
QY	141	GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp	160
DB	115188	GGTCCGGTTACGCTCTGTGACAGTCGGGCAATGTGGCGGGCTGCTGATGTTCCGAT	115129
QY	161	GlyIysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgIleu	180
DB	115128	GGCAAAATCCCGGAAAAATGCTCGAAACCGCGTATCGGGGAACTCGGATTTGGGTTG	115069
QY	181	SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn	200
DB	115068	AGCAGCGCAATCTGGAAAACTCAGCTCCGGCGCATCAATATCGTTTGGGCAAAAAAC	115009
QY	201	PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr	220
DB	115008	TTTGTATTGCAACACGAAAGGGCTGTACCGCAATCGGGGATTAACCGGTACCGCGTTAC	114949
QY	221	ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla	240
DB	114948	CGCAATCTGAAACGGCTGCCCGACAG-CCACCGCGATTTCGAAACGGGCGCATCGGGCT	114890
QY	241	ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr	260
DB	114889	GTCTGGGTGGCGAAAAAGGCTTTATCGCGCAGCATACACGACCGCTCGCGACCAATA	114830
QY	260	rClyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys	280
DB	114829	TGGTCTGCTGCCACAGCCAGCAATACGATGATTGCCACGCGGACATCATCTGGCAAAA	114770
QY	280	AspLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa	300
DB	114769	GAGTTTGATTAAACAAACCGCTATTTCGAGCTTTATCCGACCTGTTGACCGAAGAAGACAT	114710
QY	300	IaspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
DB	114709	CGATTACGACAAATCCGGGCTTGAGCTGGCGCTTTCAACGACGACGATGATGACACGCCCA	114650
QY	320	sAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG1	340
DB	114649	TGCCCAACACGGCAAACTTGGATAGACTTGGCAACAAACGCTACGAACTCCGGCGCGA	114590
QY	340	uTrrPlysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr	360
DB	114589	ATGGAAGCAACCGTTTCCCGGTTTTTGAAGCCCTGCGGTACACCTGAACCGCAACGACTA	114530
QY	360	rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs	380
DB	114529	CGCCACACGCAAAAAAGCAGCGATGTCAGTAGAAACTTTTTTAAACCAACCAACGCAAAA	114470
QY	380	nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValG1	400
DB	114469	CGCCCGTATCGATTGCGGCCCAACCATAGGCCGTCTGAAAGCGAGCTGGGGGTGCA	114410
QY	400	nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe	420
DB	114409	ATATTGGGACAAAAAATCCAGTGTCTTATCTGCCACATCCGAAGCGGTCAACACACCGAT	114350
QY	420	tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs	440
DB	114349	GCTGCTTGACAAATAAAGTGCAACATTACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGA	114290
QY	440	pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs	460
DB	114289	CAACTTCAGCTTGAAGCGGGGTACGCTGGGAAAAACAAAAAGGCTCCATCCGTACCA	114230
QY	460	pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi	480
DB	114229	CAAGCAATTGATTCGGGAAAACTACTCAACCATCCCTGCCGACCTCCGGCGCGCA	114170
QY	480	sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy	500
DB	114169	CGCCCAAAACCGCCCGCTCATTCGCACTTTCGGGCAACTGGTATTTCACGCCCAACACAA	114110
QY	500	sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi	520
DB	114109	ACTCAGCTTGACCGCTCCCATCAGAACGCTGCGGTCAACGCAAGAGCTGTACGCCACA	114050
QY	520	sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr	540
DB	114049	CGGCAAAACACGTCGCGCACCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG	113990
QY	540	sSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl	560
DB	113989	TTCCAAACAAATATCGAACTTCGGCTGCGGTACGAGCGCACCGCTGGCAATACAAATTCGC	113930
QY	560	aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG1	580
DB	113929	ACTCTACCGCAACCGCTTCGGCACTACATTACGCCCAACCTTTAAGCAGCAGCGCGG	113870
QY	580	yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl	600
DB	113869	CCCCAAATCCATCGAAGACGACAGCGAAATGAAGCTCGGTGCGCTACAACTCCGGTGC	113810
QY	600	aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG1	620
DB	113809	GGACTTCTACGGCGCGGAAGCGAATCTACTTTCAACCGCACCGCGCTACCGCATCGG	113750
QY	620	yValSerGlyAspTyrValArgGlyArgLysAsnLeuProSerLeuProGlyArgG1	640
DB	113749	CGTTTCCGGCGACTATGTACGAGCGCGCTCGAAACCTTCCTCCCTACCGCGCAGGGA	113690
QY	640	uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr	660
DB	113689	AGACGCTTACGGCAACCGCCCACTCATTTGCCCAAGCCCAACCAACCGCTCCGCTCC	113630
QY	660	oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe	680
DB	113629	GGCTGGCGCGCTTCGGGTCCACCTTGAAAGCTTCGCTGACCGCGCATCGATGCCAATTT	113570
QY	680	uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG1	700
DB	113569	GGACTACTACCGGTGTTCCGCCCAACAACTCGCCCGCTACGAAACCGCGCACGCCCGG	113510
QY	700	yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrpAs	720

Db 113509 ACACATATGCTCAACCTCGCGCGAACTACCGCGCAATACCGCTATGGCGAGTGAA 113450
 Qy 720 nTrrTyrValIysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 113449 TTGTTACGTCAAGCGCAACCTGCTCAACCAATCCGTTAGCGCCACAGCAGCTTCT 113390
 Qy 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValIysPhe 758
 Db 113389 CTCTGATACGCGCAATGGCGCGAGCTTTACCGCGCGCGTGAACGTGAAGTTT 113335

RESULT 4
 AX300375
 LOCUS AX300375 2277 bp DNA linear PAT 30-NOV-2001
 DEFINITION Sequence 201 from Patent WO0185772.
 ACCESSION AX300375
 VERSION AX300375.1 GI:17381766

KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis
 Neisseria meningitidis
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.

REFERENCE
 AUTHORS Tang, C. U.
 TITLE Virulence genes, proteins, and their use
 JOURNAL Patent: WO 0185772-A 201 15-NOV-2001;
 Microscience Limited (GB)

FEATURES
 source
 Location/Qualifiers
 1..2277
 /organism="Neisseria meningitidis"
 /mol_type="genomic DNA"
 /db_xref="taxon:487"
 1..2277
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD13141.1"
 /db_xref="GI:17381767"
 /translation="MAQTLLKPIVLISILLINTPLLAQHETESQVDLETVSVVGKSRP
 RRTVGLHSTASDKISGDTLRQKAVNLGDALGVPGIHASQYGGGASAPVIRGOTG
 RRIVLNHHGSETDMAPSPDHAIMVDTLSQVEILIRPVTLLYSSGNVAGLVDVAD
 KGIPEKPPENGSGELGIRLSGNLEKLTSGGINIGLKNFVLTHTGILYRSGDYAVP
 RYRNLKPLSHADSQTSGISLWVGKGFIVAVSDRDQYGLPASHSHVEDDCHADI
 IWQSLINKRYLOIYPLHLLTRDIDYDNPGLSCGFHDDDNAAHTHSGRPWIDLRNKR
 YELAEWKQPPGFEALREVLNRDYNRDEKAGDAVENFNNQTNARIEIURHQPIGR
 LKGSWQVYLOQKSALSASIAEVKQPMLLDNKVQHSYFFGVQANWMDNFTLEGVRY
 EKQKASIQYDKALIDRENYNHPDLPDGAHRQTARSFALSGWYFTFOHKLSTASHQ
 ERLPSTQELYAHGHKVAINTPEVGNKHLNKRSSNNIELALGYEDRWQYNLALVNRN
 GNYIYAOTLNDGRGPKSTEDDSMKLVRYNOSGADFYGAEGEIYFKPTPRYRIGVSG
 YVGRGLNPLSPGREDAYGNRPFTIADDDNAPRPAARLGFHLKASLITRIDANLDY
 YRVFAONKLARVETRTPEGHMLNLGANYRNRTRIGENWNYVKADNLLNQSVIAHSSFL
 SDTPQMGSRSTGGVNVKF"

BASE COUNT 600 a 712 c 546 g 419 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7, 28e-229 Length: 2277
 Score: 3927.00 Matches: 743
 Percent Similarity: 98.81% Conservative: 7
 Best Local Similarity: 97.89% Mismatches: 8
 Query Match: 97.30% Indels: 2
 DB: 6 Gaps: 0

US-09-936-377-2 (1-758) x AX300375 (1-2277)

Qy 1 MetAlaGlnThrThrLeuIysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
 Db 1 ATGSCACAACTACACTCAACACCATGTTTATCAATCTTTTAAATCAACACACCCCTC 60
 Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValValGly 40
 Db 61 CTCGCCCAAGCGCATGAACCTGAGCAATCGGTGGATTGTGAAACCGGTGAGCGTCTGGC 120

Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspIysIle 60
 Db 121 AAAAGCCGTCGCGCGCACCTCGCGGCTGTGGCACACTTCGACCGCTCCGACAAAATC 180
 Qy 61 IleSerGlyAspThrLeuArgGlnIysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 181 ATCTCCGCGGATACCTTCGCGCAAAAGCCGTCAACTTGGGCGACGCTTTAGACGCGTA 240
 Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 241 CCGGCGCATCCACGCTTCGCAATACGGCGCGCGCTGTGCTCCCGTCATTCGCGGTCAA 300
 Qy 101 ThrGlyArgArgIleLysValIleuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 301 ACAGCGAGCGGATTTAAAGTGTGAACCATCACGCGGAAACAGGCGGATATGCGGATTTT 360
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140
 Db 361 TCGCCCGATCAGCCATTTATGTAGATACCGCTTGTGCGAACAGGTGCAATCTCTGCG 420
 Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 421 GGGCGGTTTACGCTCTTGTACAGCTCGGCGCAATGTGGCGGGCTGCTGTCGATGTTGCCGAT 480
 Qy 161 GlyIysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 481 GGCAAAATCCCCGAAAAAATGCTGAAAACGCGTATCGGCGGAACTCGGATTCGCTTTG 540
 Qy 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 541 AGCAGCGCAATCTGGAAAAAATCATCTCCGCGCGCATCAATATCGGTTTGGGCAAAAC 600
 Qy 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 601 TTTGTATTGCACACGGAAGGCTGTACGCGAAATCGGCGGATTACGCCGTACCGCGTTAC 660
 Qy 221 ArgAsnLeuLysArgLeuLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 661 CGCAATCTGAAACGCTCGCCGCGAG-CCACGCGGATTTCGCAAAACGGCGAGCATCGGCT 719
 Qy 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
 Db 720 GTCTTGGTTGCGCAAAAGGTTTTATCGGCTAGCGTACGCGCGCGTTCGCGCAACAATA 779
 Qy 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrrGlnly 280
 Db 780 TGGTCTGCTCCCCACACCGACGATACGATGATGCCACGCCGACATCATCTGCGCAAAA 839
 Qy 280 sSerIleuIleAsnLysArgTyrLeuGlnLeuTyrProHisIleuLeuThrGluGluAspVa 300
 Db 840 GAGCTTGATTAAACACGCTATTACAGCTTTTATCCGCGACCTGTGTGACCGAAGAAGACAT 899
 Qy 300 LasPtyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
 Db 900 CGATTACGCAATTCGGGCTTGAGCTGCGGCTTCACGACGACGATAATGACACGCGACA 959
 Qy 320 sAlaHisAsnGlyLysProIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGln 340
 Db 960 CACCCACAGCGCGCAGACCGTGGATAGACCTCGCAACCAACACGCTACGAACTCCGTCGCGA 1019
 Qy 340 uTrrPlyGlnProPheProGlyPheGluAlaLeuArgValHisIleuAsnArgAsnAspTy 360
 Db 1020 ATGGAACGACCGTTCCCGGTTTTTGAAGCCCTCGCGGTACACCTGAACCGCAACGACTA 1079
 Qy 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnGlnThrGlnAs 380
 Db 1080 CCGCCACGACGAAAAAGCAGCGCATGCGTCAAAATCTTTTAAACCAACCAACGCAAAA 1139
 Qy 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrrGlyValGln 400
 Db 1140 CGCCCGCATCGAGTTGCGCACCAACCCATAGTGTGCTGAAAGGCGACGCTGGGCGTGCA 1199
 Qy 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420

```

||||| 1200 ATATTACAAACAAAATCCAGTCTTTATCTGCATATCCGAAGCGGTAAACACCGAT 1259
||||| 420 tleuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAs 440
||||| 1260 GCTGCTTGACAAAGAGTGAACATTACAGCTTTTCGGTGTAGAACAGGCAAACTGGGA 1319
||||| 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
||||| 1320 CAACCTTCAGCGCTTGAGAGGCGGTACGCGTGGAAAAACAAAAGAGCTCCATTCCAGTACGA 1379
||||| 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
||||| 1380 CAAAGCAATGATTGATCGGAAACACTACTCAACACCCCTGCCCGACCTCGCGCGCA 1439
||||| 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLy 500
||||| 1440 CCGCCAAACCGCGCTCATTCGCACCTTCGGCACTCGGTATTTACAGCCACACACAA 1499
||||| 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
||||| 1500 ACTCAGCGCTGACGCGCTCCCATCAGGAACGCGTCCGCTCAACGCAAGAGCTGTACGCACA 1559
||||| 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
||||| 1560 CGGCAAAACAGTCGCCACCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG 1619
||||| 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
||||| 1620 TTCCAACAATATGAACTCGCGCTGGCTACGAAGGCGCGCTGGCAATACATCTGGC 1679
||||| 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
||||| 1680 ACTCTACCGCAACGCGCTTCGGTAACATATTACGCCCAACCTTAAACGACGACGCGG 1739
||||| 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
||||| 1740 CCCCNAATCCATCGAAGACGACGAGCAATGAAGCTCGTGGCTACCAACCAATCCGCGCG 1799
||||| 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl 620
||||| 1800 CGACTTCTACGGCGGGAAGCGGAAATCTACTTCAAAACCGACACCGCGCTACCGCATCGG 1859
||||| 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGl 640
||||| 1860 CGTTTCGGCGGACTATGTACGAGCGCGTCTGAAACCACTCCCTTCCCTACCCGCGCAGAGA 1919
||||| 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
||||| 1920 AGATGCTACGGCAACCGTCTTTCATCGCACAGGACGACCAAAATGCCCCCGCTGTTC 1979
||||| 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
||||| 1980 GGCTGCGCGCTCGGCTTCACCTGAAAGCGCTCGCTGACCGACCGTATCGATGCCAATTT 2039
||||| 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
||||| 2040 GGACTACTACCGCGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGCAGCGCCGG 2099
||||| 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAs 720
||||| 2100 ACACCATATGCTCAACCTCGCGCAAACTACCGCGCGCAATACCGCTATCGCGAGTGGAA 2159
||||| 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
||||| 2160 TTGGTAGCTCAAAAGCCGACAACTGCTCAACCAATCCGTTTACGCCCCACAGCAGCTTCT 2219
||||| 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
||||| 2220 CTCTGATACCGCGCAAAATGGCCGCGACCTTTACCGCGCGCGTGAACGTGAAGTTT 2274

```

RESULT 5
AE002448/c

LOCUS	AE002448	10700 bp	DNA	linear	BCT 25-MAY-2000
DEFINITION	Neisseria meningitidis serogroup B strain MC58 section 90 of 206 of the complete genome.				
ACCESSION	AE002448 AE002098				
VERSION	AE002448.2 GI:7413446				
KEYWORDS					
SOURCE	Neisseria meningitidis MC58				
ORGANISM	Neisseria meningitidis MC58				
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.				
AUTHORS	1 (bases 1 to 10700) Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citterone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizzia, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R. and Venter, J.C.				
TITLE	Complete genome sequence of Neisseria meningitidis serogroup B strain MC58				
JOURNAL	Science 287 (5459), 1809-1815 (2000)				
MEDLINE	20175755				
PUBMED	10710307				
REFERENCE	2 (bases 1 to 10700) Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citterone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizzia, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R. and Venter, J.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA				
COMMENT	On Apr 4, 2000 this sequence version replaced gi:7226204.				
FEATURES	Location/Qualifiers				
source	1..10700 /organism="Neisseria meningitidis MC58" /mol_type="genomic DNA" /strain="MC58" /db_xref="taxon:122586" /note="serogroup: B" complement(116..2392) /gene="NMB0964" complement(116..2392) /genes="NMB0964" /notes="similar to GP:2114321 percent identity: 43.79; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="TonB-dependent receptor" /protein_id="AAF62323.1" /db_xref="GI:7413447" /translation="MAQTTLKPVLISILLINTPLLAQAHETEVSVDLETVSVVVKSRP RATGLHTSTASDKIISGDTLRQKAVNLGDALDGVGTHASQYGASAPVIRGQTG RIKVLNHHGTEGDMADFPDHALMWDTALSQQVEILRGVTLIYSSGNVAGLVVDVAD GKIPEKPNVSGVSGELGLSSNLEKLTSGGINIGLGNFVLHTGLYKRSBGDYAVP RYNLRKLPDSDHADSQTSGISLWVGEGKFIGVAYSDRRDQYGLPAHSHEYDCHADI IWOKLSINKRYLYQLLTHEEDIDYDNPGLSCGFHDNDNAHATHSRGIWIRNKR YELRAWKQPFQFGFEALRYHLNRNDYHDEKAGDAVENFNQNTQARIELRHOPIGR LKGSMGVLYLQOKSSALSALSAVKPMLLDNKVCHYSFPGVQAWNDNTELEGVRY EQKASIQVDKALIDRENYNPLPDGHRQTRFARSFALSGNWNFTTPQHKLSLTASHQ EKLPSTQELYAHGKHVATNFEVGNKHLNKRNSNLTALGVEYEDGRWQYNLALYRNF GNYIYAQTLNDGCGPKSIDDSMKLVRYNQSGADFYGAEGYIFKPTFRIDYVSGD YVRGLKNLPSLFRGREDAYGNRPFFAQDDQNAIPVPAARIGFLKSLASTRIDANLDY YVFAONKLARYETRTPGHMLNLGANYRNTYRGWVYVVKADNLLNQSVYHASFLL SDTPQMGSRFTGTVNKF" complement(2450..2722) /gene="NMB0965"				
gene					
CDS					
gene					

```
CDS      complement(2450..2722)
/gene="NM80965"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41370.1"
/db_xref="GI:7226205"
/translation="MRPREGAEPVUTSTPFCHLFDPSGTCTFYALINFGYPAAEIN
GQTNLRQHPQPRRTADTNGHRMPQPFHLKNFCLCPLSLKIL"
2814..3404
/gene="NM80966"
2814..3404
/gene="NM80966"
/feature="hypothetical protein"
/protein_id="AAF41371.1"
/db_xref="GI:7226206"
/translation="MLLFDIDYDSFTYNIQVYFTELGOEVAVRENDITILEIRALNP
QYLIVGPGSPKRGISVAMRHFAGRLPIMGVCLGHQTIGEAFGGRIYRAKTLMEG
KVPVSHSGKMGKGLNPVTCRYHSLVIDRNTMPECLVATWEDGEIMGVRHKEY
AVSGVQHPHALTEHGHDMNNFLIEFQNFQPKQI"
3468..4526
/gene="NM80967"
3468..4526
/gene="NM80967"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/translation="MQYVYVQSVSTGDFLYLSPETGDIGHTKLTINADYDFEEAL
NAGLEEIGNQYEFVFGFLKD"
6882..7166
/gene="NM80973"
6882..7166
/gene="NM80973"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="anthranilate phosphoribosyltransferase"
/protein_id="AAF41372.1"
/db_xref="GI:7226207"
/translation="MTPQQAIELRLISNNELFYDEMDDLMRQMRGQVLPQIAAILT
GLRKIVTSEITAAAMVREFAKVPLENAEGLDIVGTGQDGCAFNISTTSMFVA
AAAGAKVAGSGSVSSGAADVQOMGANLITPQVAGSIRQTCIGTGFMEAPNHS
AMRHVAPVRSILGRSIFNLGLPTNPAGAPNQLLGVTHTDLGILSRVLIQQLGSKHV
LVVCGEGLEILTIKTRVABKDGKISYDIRPEDFGIETRNLDIEIKVANTQSSLF
LRNVELEGREGAARDIVLENTAAALYAGNVAASLSDGISAAREADISGRAKSKREF
VGFOPOQRCHFLGKMEIG"
4534..4623
/gene="NM80968"
4534..4623
/gene="NM80968"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41373.1"
/db_xref="GI:7226208"
/translation="MIMELWQNISTRRELTKFLYKNLGLY"
complement(4678..4851)
/gene="NM80969"
complement(4678..4851)
/gene="NM80969"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41374.1"
/db_xref="GI:7226209"
/translation="MLPFIIRKKVPGDIVTDFRSYDVLDISFEFSLRKFNGIPKEH
LGLHLKCKQWHFK"
4995..6419
```

```
gene
CDS      /gene="NM80970"
/feature="This region contains an authentic frame shift and
is not the result of a sequencing artifact; conserved
hypothetical protein, authentic frameshift; identified by
Glimmer2; putative; conserved hypothetical protein,
authentic frameshift"
6382..6672
/gene="NM80971"
6382..6672
/gene="NM80971"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41375.1"
/db_xref="GI:7226210"
/translation="MNILOGHIVGVKINGIQEGKSFDCCLIVATPLDSSQNALG
SSTIYDFGGSANFEQFRNAQFFIEANLNVEITVTKTKLVIGFQLVKKG"
6677..6874
/gene="NM80972"
6677..6874
/gene="NM80972"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41376.1"
/db_xref="GI:7226211"
/translation="MQYVYVQSVSTGDFLYLSPETGDIGHTKLTINADYDFEEAL
NAGLEEIGNQYEFVFGFLKD"
6882..7166
/gene="NM80973"
6882..7166
/gene="NM80973"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41377.1"
/db_xref="GI:7226212"
/translation="MFGRLKNAPSIATAKHLKENIMKFINTCRKYGAKLAVVTAAPL
ALAAHANATIPDTAKNALEAAKADGMEAGWIVGIFAALFVSIVKRVMK"
7173..7451
/gene="NM80974"
7173..7451
/gene="NM80974"
/feature="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

Alignment Scores:
Pred. No.:      5,19e-228      Length:      10700
Score:          3927.00      Matches:      743
Percent Similarity: 98.81%      Conservative: 7
Best Local Similarity: 97.89%      Mismatches: 8
Query Match:      97.30%      Indels:      2
DB:              1          Gaps:      0

US-09-936-377-2 (1-758) x AE002448 (1-10700)

Qy      1 MetAlaGlnThrThrLeuTySProIleValLeuSerIleLeuLeuIleAseThrProLeu 20
Db      2392 ATGCACAACTACATCAAAACCCATTTGTTTATCAATTTTAAATCAACACACCCCTC 2333
Qy      21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
Db      2332 CTGCCCAAGCGCATGAACCTGAGCACTCGTGGATTGGAAACCGTCAGCGTCGTGGC 2273
```


QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 2272 AAAAGCGCTCGCGCGCCACGTCGGGCGCTGTTGCACACATTCGACCGCTCCGCAAAATC 2213
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 2212 ATCTCGGCGATACCTTGCCGCAAAAGCGCTCACTTGGGCGACGCTTTAGACGGCGTA 2153
 QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 2152 CCGGGCATCCACGCTTCGCAATACGGCGCGCGGCTGCTCCCGTCATTCGCGGTCAA 2093
 QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 2092 ACAGGCGAGCGGATTAAGATGTTGAACCATCACGGCGAAACAGCGCATATGCGGGAATTT 2033
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140
 Db 2032 TCGCCCGATCACGCCATTATGGTAGATACCGCTTGTCGCAACAGTTCGAATTCCTGCGC 1973
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 1972 GGGCGCGTTACGCTCTGTACAGCTCGGCAATGTGGCGGGGCTGGGTGCGATGTGCGCAT 1913
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 1912 GGCAAAATCCCCGAAAAATGCTTGAAACCGCGGTATCGGGCGAATCGGATTTGCGTTG 1853
 QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 1852 AGCAGCGCAATCTGGAAAAAATCACGTCGCGGCGCATCAATATCGGTTTGGCAAAAC 1793
 QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 1792 TTTGTATTCACACGGAAGGGCTGTATCCGCAAAATCGGGGATTCACGCGTACGCGGTAC 1733
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 1732 CGCAATCTGNAACGCTGCCCGACAG-CCACGCGGATTCGCAACCGGCGAGCATCGGGCT 1674
 QY 241 ValLeuGlyTrpArgLysArgPheTyrArgArgThr-TyrSerAspArgAspGlnTyr 260
 Db 1673 GTCTTGGTGTGGCAAAAGGTTTATCGGCGTAGCGTACAGCGACGCTCGCGACCAATA 1614
 QY 260 rGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTrpGlnLys 280
 Db 1613 TGTCTCGCTGCCCAACAGCCAAATACGATGATTCACGCGGACATCATCTCGGCAAA 1554
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
 Db 1553 GAGCTTGATTACAAACGCTATTATACAGCTTTATCGGACCTGTTGACCGAAGAGACAT 1494
 QY 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
 Db 1493 CGATTACGAAATCCGGGCTTGAGCTGGCGTTCACAGCAGCATATAATGACACGCA 1434
 QY 320 sAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG 340
 Db 1433 CACCCACGGCGGACCGGTGAATAGACCTTGGCGCAACAAACGCTACGAACTCGTGCCGA 1374
 QY 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
 Db 1373 ATGGAAGCAACCGTTCCTCCCGGTTTGAAGCCCTGCGGTACACTGAAACCGCAACGACTA 1314
 QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 Db 1313 CCGCCACAGCAAAAGCAGCGCATGCGATCGAAAACTTTTAAACCAACACGCAAAA 1254
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValG 400
 Db 1253 CGCCCCGATCGAGTTGCGGCCACCAACCCATGATGCTGTGAAGGAGCGCTGGGGCGTGCA 1194
 QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProWe 420

RESULT 6
 AX044031/c

Db 1193 ATATTTTCAACAAAAATCCAGTGTCTTTATCTGCGATATCCGAAGCGGTTAAACCAACCGAT 1134
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 Db 1133 GCTGCTTGACAAAGATGCAACATTACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGA 1074
 QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 Db 1073 CAACTTCACGCTTGAAGGAGCGGTACGCTGGGAAAAACAAAGAGCTTCCATTTCAGTACGA 1014
 QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProLeuProAspLeuGlyAlaHi 480
 Db 1013 CAAAGCATTTGATTGATCGGAAAAACTACTACAACCAACCCCTCCCGACCTCCGCGCGCA 954
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLys 500
 Db 953 CCGCCAAAACCGCCCGCTCATTCGCACTTTCGGGCAACTGGTATTTTCAAGCAACACAA 894
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 Db 893 ACTCAGCTGACCGCTCCCATCAGGAACGCTTCGCTCAACGCAAGAGCTGTACGCACA 834
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 Db 833 CGCAAAACAGCTCGCCCAACCACTTTGAAGTCGGCAACAAACCACTCAACAAAGAGCG 774
 QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
 Db 773 TTCCACAATATCGAATCTCGCTGGCTACGAAGGCGACCGCTGGCAATACAAATCTGGC 714
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG 580
 Db 713 ACTCTACCGCAACCGCTTCGGTAACTACATTTACGCCCAAACTTAAACGACGACGCGG 654
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 Db 653 CCCCAATCCATCGAAGACGACGCAAAATGAAGTCGTCGCTTACCAACCAATCCCGCGC 594
 QY 600 aAspPheTyrGlyValaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG 620
 Db 593 CGACTTCTACCGCGCGGAGGCGAATCTACTTCAACCGACACCGCGCTACCGCATCGG 534
 QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG 640
 Db 533 CGTTTCCGCGCACTATGTACGAGCGCTCTGAAAAACCTGCCTTCCCTACCCGCGCAGAG 474
 QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 Db 473 AGATGCTTACCGCAACCGCTTTCATCGCACAGGACGACCAAAATGCCCCCGCTGTTC 414
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
 Db 413 GGTGCGCGCTCGGCTTCCACTGAAAGCTCGCTGACCGACCGTATCGATGCCAATTT 354
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
 Db 353 GGACTACTACCGCGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGACGCGCGG 294
 QY 700 yHisHisMetLeuAsnLeuGlyValaAsnTyrArgArgAsnThrArgTyrGlyuTrpAs 720
 Db 293 ACACCATATGCTCAACTCGGCGCAAACTACCGCCCAATACGCGCTATGCGCGAGTGGAA 234
 QY 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 233 TTGTAGCTCAAGCGCAACCTGCTCAACCAATCCGTTTACGCCCAACGACGACTTTCT 174
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 173 CTCTGATACGCGCAAAATGGCGCGAGCTTTTACCGGCGCGGTGAACGTCGAAGTTT 119

LOCUS AX044031 349980 bp DNA linear PAT 24-NOV-2000
 DEFINITION Sequence 110 from Patent WO0066791.
 ACCESSION AX044031
 VERSION AX044031.1 GI:11342915
 KEYWORDS
 ORGANISM Neisseria meningitidis
 Neisseria meningitidis
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE 1
 AUTHORS Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,
 Maignani, V., Galeotti, C., Mora, M., Ratti, G., Scariselli, M.,
 Scarlato, V., Rappuoli, R., Frazer, C. M. and Grandi, G.
 TITLE Neisseria genomic sequences and methods of their use
 JOURNAL Patent: WO 0066791-A 110 09-NOV-2000;
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
 FEATURES
 source
 1. 349980
 /organism="Neisseria meningitidis"
 /mol_type="genomic DNA"
 /db_xref="taxon:487"
 /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
 349980 349980 bases-seq 108: 300001 to 649980 349980
 bases-seq 109: 600001 to 949980 349980 bases-seq 110:
 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
 2272325 172325 bases"
 BASE COUNT 81351 a 86755 c 95584 g 86290 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.34e-226 Length: 349980
 Score: 3927.00 Matches: 743
 Percent Similarity: 98.81% Conservative: 7
 Best Local Similarity: 97.89% Mismatches: 8
 Query Match: 97.30% Indels: 2
 DB: 6 Gaps: 0
 US-09-936-377-2 (1-758) x AX044031 (1-349980)
 Qy 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAasnThrProLeu 20
 Db 85503 ATGGCAAACTACACTCAAAACCATGTTTATCAATTCCTTTTAAATCAACACACCCCTC 85444
 Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 Db 85443 CTCGCCAAGCGCATGAACACTGAGCAATCGGTGGATTGGAAACCGGTGAGCGTCTCGGC 85384
 Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 85383 AAAAGCGTCCGCGCCACGTCGGGCTGTTCACACTTCGACCGCTCCGACAAATC 85324
 Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 85323 ATCTCCGGCGATACCTTTCGCGCAAAACCGCTCAACTTGGCGACGCTTTAGACGGCGTA 85264
 Qy 81 ProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 85263 CCGGGCATCCAGCTTCGCAATACGGCGCGCGCTGCTGCTCCCGTCATTTCGCGGTCAA 85204
 Qy 101 ThrGlyArgGlyIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 85203 ACAGGAGCGGATTAAGGTGTAACCATCACGCGCAACACGCGATATGGCGGATTTT 85144
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 85143 TCGCCCGATCACGCCATTATGTAGATACCGCTTGTTCGCAACAGGTGCAATCTCTCGC 85084
 Qy 141 GlyProValThrLeuLeuThrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 85083 GGGCCGGTTACGCTCTTGTACAGCTCGGGCAATGTGGCGGGCTGGTTCGATGTGGCGAT 85024

Qy 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 85023 GGCAAAATCCCGGAAAAAATCCCTGAAACACGCGTATCGGGCGAACTCGGATTTGGTTTG 84964
 Qy 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 84963 AGCAGCGGCATCTGGAAAAAACTCAGTCCGCGCGGCATCAATATCGGTTTGGGCAAAAC 84904
 Qy 201 PheValLeuHisThrGluGlyLeuThrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 84903 TTTGTATTGCACACGGAAGGCTGTACCGCAATTCGGGGGATTACCGCGTACCGGTTTAC 84844
 Qy 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 84843 CGCAATCTGAACACGCTGCCGCGACAG-CCACGCGGATTCGCAACCGGCAGCATCGGCT 84785
 Qy 241 ValLeuGlyTyrArgLysArgPheThrArgArgThr-TyrSerAspArgAspGlnTyr 260
 Db 84784 GTCTTGGGTTGGCGAAAAAGSTTTTATCGGCGTAGCTACAGCGACCGTCGCGCAATA 84725
 Qy 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
 Db 84724 TGGTCTGCTGCCACAGCCACGAAATACGATGATTGCCACCGCACATCATCTGCAGAAA 84665
 Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
 Db 84664 GAGCTTGATTAAACAAACGCTATTTTACAGCTTTATCCGCACTGTGTGACCGAAGAGACAT 84605
 Qy 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
 Db 84604 CGATTACGACNATCCGGGCTTGAGCTGGGCTTCCAGACGACGATATGTCACACGCA 84545
 Qy 320 sAlaHisAsnGlyLysProThrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGly 340
 Db 84544 CACCCACAGCGGCAGACCGGTGATAGACCTCGCGCAACAAACGCTACGAACTCCGTCGCA 84485
 Qy 340 uTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
 Db 84484 ATGGAAGCAACCGTTCCCGGTTTTTGAAGCCCTTCGCGCTACACCTGACACCGCAACGACTA 84425
 Qy 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnGlnThrGlnAs 380
 Db 84424 CCGCCACGACGAAAGACGCGCATGACGTGCAAACTTTTTTAAACACCAACGCAAAA 84365
 Qy 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGly 400
 Db 84364 CGCCCGCATCGAGTTGCGCACCAACCATAGTCTGCTGAAAGCGAGCTGGGCGTGCA 84305
 Qy 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
 Db 84304 ATATTTACAAACAAAATCCAGTGTCTTATCTGCCATATATCCGAAGCGGTAAACACCCGAT 84245
 Qy 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnThrAs 440
 Db 84244 GCTGCTTGACAAACAAAGTGCAACATTCAGCTTTTCGTGTAGAACAGGCAAACTGGGA 84185
 Qy 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 Db 84184 CAACCTCAGCGTTGAAGAGGCGGTACGCGTGGAAAAACAAAACCCCTCCATTCAGTACGA 84125
 Qy 460 pLysAlaLeuIleAspArgLysThrLysGlnProLeuProAspLeuGlyAlaHis 480
 Db 84124 CAAAGCATGATGATCGGAAAACTACTACAAACCCCTCCCTGCGCGCGGCA 84065
 Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnThrTyrPheThrProGlnHisLys 500
 Db 84064 CCGCCAAACCGCGCTCATTCGACCTTCGGGCACTGGTATTTTCAGCCACCAACACAA 84005
 Qy 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 Db 84004 ACTCAGCCTGACCGCTCCCATCAGGAACGCTCCGCTCAACGCAAGAGCTGTACGCACA 83945
 Qy 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540

Db	83944	CGGCAACACGTCGCGCACCAACACCTTGAAGTCGCGACACAAACACCTCAACAAAGAGCG	83885
Qy	540	gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl	560
Db	83884	TTCCACCAATATCGAACTCGCGCTCGGCTACGAGCGACCGCTGCAATACAAATCTGCG	83825
Qy	560	aleuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl	580
Db	83824	ACTCTACCGCAACCGCTTCGGAACCTACATTTACGCCCAACCTTAAACGACGAGCGCGG	83765
Qy	580	yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl	600
Db	83764	CCCCAAATCCATCGAAGACGACAGCAATGAAGCTCGTGCCTACAAACATCCGCGCG	83705
Qy	600	aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl	620
Db	83704	CGACTTCTACGCGCGGAGGAGCGAATCTACTTCAAAACGACACCGCGCTACCGCATCGG	83645
Qy	620	yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGl	640
Db	83644	CGTTTCCGCGGACTATGTACGAGGCGCTCTGAAAACCTGCTTCCCTACCGCGCAGAGA	83585
Qy	640	uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr	660
Db	83584	AGATGCTTACGCGCAACCGCTCTTCATCGCACAGCAGCAACAAATGCCCGCGTTTCC	83525
Qy	660	calaalaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe	680
Db	83524	GCTCGCGCGCTCGGCTTCCACTGAAAGCCTCGCTGACCGACCGATCGATGCCAATTT	83465
Qy	680	uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl	700
Db	83464	GGACTACTACCGCGTTTCGCCCAAAACAACTCGCCCGCTACGAAACCGCGCCCGG	83405
Qy	700	yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLTrpAs	720
Db	83404	ACACCATATGCTCAACCTCGCGGCAAACTACCGCGCAATACCGCGCTATGCGGAGTGGAA	83345
Qy	720	nTrpTyrValLysAlaAspAsnLeuLysGlnSerValTyrAlaHisSerSerPheIle	740
Db	83344	TTGGTACGTCAAGCGGCAACCTGCTCAACCAATCCGTTTACGCCCAACGAGCTTTCT	83285
Qy	740	uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe	758
Db	83284	CTCTGATACGCGCAAAATGGCGCGAGCTTTACCGCGCGGAGTGAAGCTGTTT	83230
RESULT 7	AY028475	4931 bp DNA linear BCT 29-JAN-2002	
LOCUS	AY028475	Mannheimia haemolytica ribonuclease PH-like protein and	
DEFINITION	AY028475	iron-regulated outer membrane protein (irp) genes, complete cds.	
ACCESSION	AY028475.1	GI:13591378	
VERSION	AY028475.1	GI:13591378	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			

source	1. .4931	/organism="Mannheimia haemolytica"	
CDS		/mol_type="genomic DNA"	
		/db_xref="taxon:75985"	
	1063..1779	/codon_start=1	
		/evidence=not experimental	
		/transl_table=11	
		/product="ribonuclease PH-like protein"	
		/protein_id="AAK29744.1"	
		/db_xref="GI:13591380"	
		/translation="MRPNRNANNOVREIKITNTRYAEGSVLIEFGTEKVLGNATIVE	
		ESVRFPLGQGGWTAAYGMLPRATHSTOREAAKGGQGGRTMEIQILASLSRAVV	
		ALKALGERTVDCVIOADGGRTRPASTIGACVALLYDANKLIVANGTILKTINPMKGLVA	
		DISGVIVAGEAVCDLEYLEDNSAETDMNVVMVQDGLRVEVQGTAEGETPFSHOELLTL	
		DLAKGTSQLYEAQKALAA"	
gene	2003..4303	/gene="irp"	
CDS	2003..4303	/gene="irp"	
		/function="involved in iron acquisition"	
		/note="TonB-dependent"	
		/codon_start=1	
		/transl_table=11	
		/product="iron-regulated outer membrane protein"	
		/protein_id="AAK29743.1"	
		/db_xref="GI:13591379"	
		/translation="MLKKNYLTVSILLAISGVYANEISLETITVDGNTPTGKLLG	
		GELNSVSDENKLGKQSTITLGNALSGELGTHSQFGGASTPIRQGESKRAKILQ	
		NNGNLDMGSMPOHVAVTVDALLAKRIELRPGPTLLYSAGNTAGVINVDNKIPTAI	
		PEKVGEGFQVRFSGSASERLTYSAGTSTFALGNHLALRVQGMKASVYAPHTIEGK	
		PYHRVPSDVQSQGTGTSLSWIGERHGLGIATDRDXKGLIHTHKKDHYTISIIRQ	
		AVMFAGYLRTFPLAEEGDIDYNNPGLIRLHHPGGSHYQGDTHHGKPMWHEK	
		RYDIDGSLQNPFGFEAKISANYVDYHDEKQKRVENFKNGKNLRFELVHKWK	
		GLKGAIGVYTNQSTSALEASRAAKVFNKPLNNPKTLWSLFAIRLNLGDTTF	
		ELSGRAERQKIAMDYDKLIDRWLGFTNPMPLNDPHKDKGYSYSEATHWYFAPNKL	
		LNAHOBRLPNAOELYAHGKIALNAPEAGNKNLKKERSNOIELSLAYVGDKWDYKLN	
		LYHRYGNYIPLTILNDNRGPKSFDEYNLKNRYQQEARSFGSAGEGIGLYLTNNYR	
		LAVGDYVRGLWNLNPMNSYNINWIGVDKASQDIPRIPPLRIGARNAQDFNL	
		NWGSLEYRYRFAQKVKYQVPTVPHHQNVLGVTSYHNFQTEYQVFLKVDNLLNQK	
		MYQASLYLPHIPQMGKNAMLGWNTSF"	
BASE COUNT	1519 a 1022 c 957 g 1433 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	4..02e-95	Length:	4931
Score:	1725.50	Matches:	346
Percent Similarity:	60.61%	Conservative:	128
Best Local Similarity:	44.25%	Mismatches:	274
Query Match:	42.75%	Indels:	35
DB:	1	Gaps:	8
US-09-936-377-2 (1-758) x AY028475 (1-4931)			
Qy	2	AlaGlnThrThrLeuLysProIleLeuLeuSerIleLeuLeuIleAsnThrProLeuLeu 21	
Db	1985	TCTAGAGGACTCTAAACGCTGTTAAGAAAATATCTCAGTATCAATTCCTCTGCC 2044	
Qy	22	AlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys 41	
Db	2045	ATTAGTGAGTAGGCTACGCTAATGAAATCTCTTTAGAGACTATTACTGTTGATGAAAT 2104	
Qy	42	SerArgProArgAlaThrSerGlyLeuLeuHisThrSer---ThrAlaSerAspLysIle 60	
Db	2105	ACG---CCTCAACCAAGGGAATTTACTAGTGTGTGAACATAAATGAATCTGTT 2161	
Qy	61	IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80	
Db	2162	GTTGATGAAAAAATTAAACAAGGTAGTATTACATTAGGTAAACGCTTTATCTGGTGA 2221	
Qy	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100	
Db	2222	TTAGGCATTTCATCAAGCCAGTTCGGAGCGGTGCAAGTACCCCAATATTTCGTGGCAA 2281	


```

VERSION      AX052567.1  GI:12226784
KEYWORDS
SOURCE
ORGANISM      Moraxella catarrhalis
               Moraxella catarrhalis
               Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
               Moraxellaceae; Moraxella.
REFERENCE
AUTHORS      1. Thonnard, J. S.
TITLE        Novel compounds from moraxella catarrhalis
JOURNAL      Patent: WO 0071724-A 9 30-NOV-2000;
              SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
source       1..2457
              /organism="Moraxella catarrhalis"
              /mol_type="genomic DNA"
              /db_xref="taxon:480"
BASE COUNT   744 a ~560 c 554 g 599 t
ORIGIN
Alignment Scores:
Pred. No.:   5,35e-87      Length: 2457
Score:       1585.00      Matches: 338
Percent Similarity: 58.92% Conservative: 141
Best Local Similarity: 41.57% Mismatches: 272
Query Match: 39.27% Indels: 63
DB:          6 Gaps: 13

US-09-936-377-2 (1-758) x AX052567 (1-2457)
QY      7  LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db      22 AAACCGCTGCTGCGCATATGGCACTTTTCAATGCAATGCTGGCAGAGGCGAAT 81
QY      26 ---GluThrGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
Db      82 TTAAGGATAAGCAACCGTCATTTAGATGGCGTTTCGATCACCTCTTTAGTGACCAA 141
QY      45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db      142 AATACAGAGTTTGGCGTTAATCATCTCAAAACACAGTCAGTGGCATCAAGTTTCAAAAGAG 201
QY      65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db      202 CAACTACACACAGGACCAACCCCTAGGCGATGCTTGGCAGTGAGCTTGGCGTTCAT 261
QY      85 AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
Db      262 TCTAACCAATTTGGGGCGGTGCTCAGCCCCCATCATTCGTGGCAGGAGGTAAACGC 321
QY      105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db      322 CTGAATCTCAAAACCGTTGAGAGTTGTGGACATGTCTGGTTGTGCGCAGACCAT 381
QY      125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
Db      382 GCCATAGCGTGGACACCACTGGCAAAACAGGTGGAGATTGTGGAGGCTCTGGTGC 441
QY      145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db      442 TTGTGTACCCCTCTGGCACTCAGCAGGCGTGTCAATGCTGTTGATGACAAATACCC 501
QY      165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db      502 AGCAATITGCCACGAAA---TTACAGGTGATGTGACGGTGTGCTTTAGCATGGCAAC 558
QY      185 LeuGluLysLeuThrSerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db      559 CGTGAATAATATCACCGCCAGTGGCGAAGCCCACTGGGAGAGCATGTGGCAGTGCCT 618
QY      205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db      619 GTTGCAGGGTGTCCAAACAGGACGACGACTATAAAACGCCAGCGCTTTGACCGCCATGTC 678

```

```

QY      221 -----ArgAsnLeu 223
Db      679 TTTAACAACAAACATGAAGATGATAACACTCAGCAGCAATTCATCTATAAGACACCTTA 738
QY      224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db      739 AAGCATCTGCCACAGCCCATGCCAATCAAAACGAGAAACGCT- TGGCGTGTGTCATGGT 797
QY      244 TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPr 263
Db      798 TGGCAATCAAGGCTTTTGGGGCATCGGTGAGCTTACGCCGAGACAAATATGGCTTGC 857
QY      263 oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuI 283
Db      858 CAACCATTCACATGAATATGAAGAATGTAGCGTGCATGGCATTTCTCAGTCCGCTTACA 917
QY      283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
Db      918 ATACAAGCCCATATTTGGCTTTGTATCTTTTGTATGGAATAATCATGACTTAGAGTTGA 977
QY      303 pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi 318
Db      978 CAATGACGGCTTGAATGCCATACATGATGATGACGACGACGACGACGACGACGACGAC 1037
QY      318 sAlaHisAlaHis-----AsnGlyLysProTrpIleAspLeuArgAsnLys 333
Db      1038 TGACCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1097
QY      333 sArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgVa 353
Db      1098 GCGTTATGATGTCAAGGCGCAATCAATGCGCGTGTGTGGCATTTGATATAATCCGAGC 1157
QY      353 lHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPh 373
Db      1158 CAGCATGGGCAGCAAGTATGATATCATCATGATGATGATGATGATGATGATGATGATGAT 1217
QY      373 ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
Db      1218 TTTTGTATAATCAAGCAATGTGTGGCGTCTCGAAGCCTCACATACCCCATTCATACGCC 1277
QY      391 ---GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
Db      1278 GATGGGCAAGTTTAGCGGGGTGTTTGGGGTAGGTTATCTCACCCTCAAAAACAGCGGACT 1337
QY      409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
Db      1338 TGTGCCACCTCGTTATGAGATGCGCAATAAACAAGACACCCCAACATCTTGCACATAA 1397
QY      424 nLysValGlnHisTyrSerPheGlyValGluGlnAlaAsn---TrpAspAsnPheTh 443
Db      1398 TAAACCAACAAACAGGCGAGTGTGTTTGGTTTGAAGAATACAAACCCCAATGACAAGCTGAC 1457
QY      443 rLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
Db      1458 CGTTGACGCGCGCTCGCATTTGAGAACAACCAATCACCATCGATTATGATAAAGACGC 1517
QY      463 uIle-----AspArgGluAsnTyrTyrLy 471
Db      1518 CATTTATCAGAGCTTAACTTAGGCTTAGCAACCGCTCATGAACGACGACATACGCTTAA 1577
QY      471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
Db      1578 ACGATTGTCTGACAGCGGTACTTTTAAACCCCAACAAACAAACCGCAGCTCTTATGCTGT 1637
QY      489 uSerGlyAsnTyrTrpPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
Db      1638 TGGGACGCAATTTACAATTACGCCCAACATAAATATTCGCTGAATCTCTCGCATCAAGA 1697
QY      509 uArgLeuProSerThrGlnLeuLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
Db      1698 ACGCTTGCCAAATGTCTCAGGAATTTGATGCTCAGCGCATGACCTTGCCACCACTCGTT 1757
QY      529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549

```

```
Db      1758  TGAATTTGGCAACCGCTTTTAAACAAAGAAAATCCACACATTTGATTTGGGCTGAC 1817
Qy      549  yTyrGluGlyAspA-gTTPGlnTyrAsnLeuAlaLeuTyrA-gAsnA-rPheGlyAsnTy 569
Db      1818  ATTTCAGGTGATAAATGGGATTCATCGTCTGGGGCTATCATATTGATTTTGATACTA 1877
Qy      569  rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerSelleGluAspSe 588
Db      1878  TGTGTTTACAAACATTTCCGAGTATAGCAAGGTTTGGCGTCATGCGTCATGATAA 1937
Qy      588  rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyG 608
Db      1938  AGACITAAACCGCAGCGCTATGAACAGCAGCGGGAATTTTGGATTTGATGCTCAA 1997
Qy      608  uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArg 628
Db      1998  CATCGGTTATCAGATTAAATGATGATATATCATGTCGGGTTATTGGTGATATATTC 2057
Qy      628  yArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPro 648
Db      2058  CAACTCTACCAATTTGCCTGACAAAAGGGCAGAACCGATGCGTATGGCAACGCTCT 2117
Qy      648  eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHis 668
Db      2118  CATCAACAGCGCAGACAGTCATAGCGCAAGACTGCCACCAAAACGCTTGGCATGAA 2177
Qy      668  uLysAlaSerLeuThrAspA-rIleAspAlaSerLeuAspTyrTyrArgValPheAla 688
Db      2178  AACCCGCAATGTTAAATGCAAAATGGTCAGGGTTTGTGGAATATCGCATACCTTT 2237
Qy      688  nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGly 708
Db      2238  AGATAAAATGGGAATTTTGACCGCCACACCCAGCTCATTAAGTGGTGAATTTGGG 2297
Qy      708  aAsnTyrArgA-gAsnThrArgTyr-----GlyGluTrpAsnTrpTyrValLysAla 726
Db      2298  TAACTATCAGCAACAGCAAGCAAGCGCTCGGTCAGGTATTTTAAATGCTAA 2357
Qy      726  pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGln 746
Db      2358  CAATCTATTAAACGATAAAGTCTTGCTCATGAGACATTTTCCACAGACATGCCACA 2417
Qy      746  tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      2418  GGGCGAAACTTATGCTCGGGCAAAATTCAAATTT 2454

RESULT 9
AX394914
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Moraxella catarrhalis
Moraxella catarrhalis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
1
REFERENCE
AUTHORS
Loomere,S., Wang,J., Bradley,B., Ochs,M. and Yang,Y.P.
TITLE
Moraxella polypeptides and corresponding dna fragments and uses
thereof
JOURNAL
Patent: WO 0218595-A 51 07-MAR-2002;
Aventis Pasteur Limited (CA)
FEATURES
source
Location/Qualifiers
1..2646
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
101..2545
/note="unamed protein product"
/codon_start=1
/transl_table=11

CDS
```

```
/protein_id="CAD32083.1"
/db_xref="GI:21065988"
/translation="MIKKPLACAILATTFSPMLAEANLKDPTVILDGVSIITSLADQN
TEFGVNSHTVSGITVSKREQLQOQRTTLGDALGELGVHSHNPFHGGASAPIIRGQEGK
RLKIQNGSEVVVMSGLSPDHAIAVDVTLAKQVEIVRSGALLYASNGAGVNVVDD
KIPSKLPKLGQDVTQVRLSSANREKLIITASAEAPLGEHVAHVAKSQAADYKTPRF
DRHFNKKHEDDNTQPEFYKDTLKLHPDSHAKSNAGTGLGVWVGNQGLGASVLR
DKYGLPNHSHEYERESVHGISQSAQYKPYLRLYPLFMDNDLLEFNNAGLECHTHDDH
DEHDHADDHDEHDEHDKWIDILMKRYDVQGINAPFAGIDKIRASMGKVDYHHD
DEHDEKTSYFDNQANVNRLEASHTPIHTPMGKSGVGVLYLTSKNSGLVPPYEDG
NQDQNTILHNKTKTSVWFPEEYKPNDKLTVDAAARIEKQITMDYQKDAIYQSLN
LGLAHABEDIRPKLLDSGTLPKPKOTARSVAVQTHLQLTTPKHKLNLNLSHDERLPN
AGELYAHGMLATNSFELGNRPLNKEKSNIDILGTFQGDWDRYLGGVHYHDFNVF
LQTLQYQKGLRMRHDKLKYARQEAAAKYGFEDVNIQYQINDYHVALFGDYIRG
KLUTNPDKKRGRTDAYGNRPILKOPDSHTPLPPLKMLGTANVNANWAGFLDYRHF
WQDLNANPERPTPAHNLNLNLGNYQHKPSHQAGSVQVFFNANLLNDRKFAHETFPFD
MPQMRNFMGLGNFKF"
BASE COUNT      817 a      583 c      571 g      675 t
ORIGIN
Alignment Scores:
Pred. No.:      5,87e-87      Length:      2646
Score:      1585.00      Matches:      338
Percent Similarity:      58.92%      Conservative:      141
Best Local Similarity:      41.57%      Mismatches:      272
Query Match:      39.27%      Indels:      63
DB:      6      Gaps:      13
US-09-936-377-2 (1-758) x AX394914 (1-2646)
Qy      7  LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db      110  AAACCGCTTGGCTGTGCCAATATGGCAACTTTTCAATGCCAATGCTGGCAGAGGCGAAT 169
Qy      26  ---GluThrGluGlnSerValGlyLeuGluThrValThrValGlyLysSerArgPro 44
Db      170  TTAAAGGATAAGCAACCGCTCATTTTAGATGGGCTTTCGATCACCTCTTTAGCTGACCAA 229
Qy      45  ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db      230  AATACAGAGTTTGGGCTTAATCATTCAAAACAGTCAGTGGCATCACAGTTTCAAAGAG 289
Qy      65  ThrLeuArgGlnLysValAlaAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db      290  CAACTACAAACAGCAGCAACCCCTAGGCGATGCTTGGCAGGTGAGTTTGGCGTTCAT 349
Qy      85  AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
Db      350  TCTAACCATTTTGGGGCGGTGCTCAGCCCCCATCATCTCGTGGCAGGAGGTAACGC 409
Qy      105  IleLysValLeuAsnHisGlyGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db      410  CTGAAATCTCTACAAAACCGTTTCAGAGTTTGGACATGTCGTGGTGTTCGCCAGACCAT 469
Qy      125  AlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGlyProValThr 144
Db      470  GCATAGCGGTGACACCACTGGCAAAACAGGTGAGATTGTGCGAGGCTCTGGTGCC 529
Qy      145  LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db      530  TTGTTGTAGCTCTGGCAACTCAGCAGCGGTGTCATGTCGTTGATGACAAATATCCC 589
Qy      165  GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db      590  AGCAAAATTTGCCCAAGCAAA---TTACAAGGTGATGTGACGGTGTCTTAGCAGTGCAC 646
Qy      185  LeuGluLysLeuThrSerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db      647  CGTGAATAATTAATCAGCCCGAGTCCGAAGCCCCCATCTGGAGAGCATGTGGCAGTGC 706
Qy      205  ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db      707  GTTGCAGGGCTGTCCAAACAGCAGCAGACTATAAACGCCACGCTTTGACCGCCATGTC 766
```

```

QY      221  -----ArgAsnLeu 223
Db      767  TTTAAACAAAAACATGAAGATGATAAACAACCTCAGCCAGAAATTCATCTATAAAGACACCTTA 826

QY      224  LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db      827  ARGATCTGCCACAGACCGCATGCCAATCAACGAGGAACGGT-TGGCGTGTTCATGGGT 885

QY      244  TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPr 263
Db      886  TGGCAATCAAGGCTTTTGGGGGCATCGGTGAGCTTACGCCGAGACAAATATGGCCTGCC 945

QY      263  oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIl 283
Db      946  CAACCAATTTCATGAATATAGAAGATGTAGCGTCATGGCAATTTCTCAGTCGCCCTTACA 1005

QY      283  eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
Db      1006  ATACAAGCCATATTTGGTTTGTATCTTTTGTATGGAAATATGATGACTTAGAGTTTGA 1065

QY      303  pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi 318
Db      1066  CAATGAGGGCTTGAATGCCATACATGATGATGACCAAGCAGCAGCAACCATGCTCA 1125

QY      318  sAlaHisAlaHis-----AsnGlyLysProTrpIleAspLeuArgAsnLys 333
Db      1126  TGACCAAGGACGACGACGAGCAGCAGCATGGCAACCTTGGATGATTTGAAATGAA 1185

QY      333  sArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgVa 353
Db      1186  GCGTTATGATGTCGAAGGCAATCAATGCGCGCTTTGTGGCATGTGATAAATCCGAGC 1245

QY      353  lHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAsnPh 373
Db      1246  CAGCATGGGCAGAAAGTGATTTATCATCATGATGATAGATAGTGGGGTGAGAAGACCAGCTT 1305

QY      373  ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
Db      1306  TTTTGATAATCAAGCAATGTGGCGTCTGGAAGCGCTCATACCCCAATTCATACGCC 1365

QY      391  ----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
Db      1366  GATGGCGCAAGTTTAGCGGGGTGTTGGGTAGGTATCTCACCTCAAAAACAGCGGACT 1425

QY      409  u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
Db      1426  TGTGCCACCTCGTTATGAAGATGGCAATAAACAAGACACCCAAAACATCTTGCACAATA 1485

QY      424  nLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn--TrpAspAsnPheTh 443
Db      1486  TAAACCAAAACAGGCGAGTGTGTTTGGTTTGAAGAATACAAACCAATGACAAAGCTGAC 1545

QY      443  rLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
Db      1546  CGTTGACCGCGCGCTCGCATTTGAGAAACAAACCAATCACCATGATATGATATAAGACGC 1605

QY      463  uIle-----AspArgGluAsnTyrTyrLy 471
Db      1606  CATTTATCAGAGCTTAAACTTAGGCTTAGCAACCGCTCATGAACCAACAGCATACGCTTAA 1665

QY      471  sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
Db      1666  ACGATTGCTGGACAGCGGTACTTTAAACCCCAAAACAAACCGCAGCTCTATGCTGT 1725

QY      489  uSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
Db      1726  TGGGAGCGCATTTACAAATTAACGCCCAACATAAATATCGCTGAATCTGTGTCATCAAGA 1785

QY      509  uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
Db      1786  ACGCTTGCCTCAATGTCTGAAGATTTGTATGCTCAGCGCATGCACTTGGCCCAACTCGTT 1845

```

```

QY      529  eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
Db      1846  TGAATTTGGCAACCGCTTTTAAACAAGAAAAATCAACAACATTTGATTGGGGCTGAC 1905

QY      549  YTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
Db      1906  ATTTCAGGGTGAATAATGGGATTCGTCTTGGGGCTATCATATTGATTTGATAACTA 1965

QY      569  rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588
Db      1966  TGTGTTTTTACAAACATTTGCGCAGTATAAGCAGGTTTTCGGTGGCATCGCTCATGATA 2025

QY      588  rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGl 608
Db      2026  AGACTTTAAACACCGCAGCTATGAACAAGCAGCGGCAAAATTTATGATTTGATGTCAA 2085

QY      608  uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl 628
Db      2086  CATCGGTTATCAGATTAAATGATGATCATGTGGCGTTATTGGTGATTTATTCGTGG 2145

QY      628  YArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
Db      2146  CRAAGCTCACCAATTTGCTTCACAAAAGGCGACGACCGATGCGTATGGCAACGCTCTCT 2205

QY      648  eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db      2206  CATCAAAACAGCCAGCAGTATACGCGCAAGACTGCCACCAAAACGCTTGGCATGAAAT 2265

QY      668  uLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGl 688
Db      2266  AACCGCAATGTAATGCAAAATGTCAGGTTTGGTAATATCGCCATACCTTTAAACA 2325

QY      688  nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708
Db      2326  AGATAAATTTGGGAAATTTGAACGCGCAACCCAGCTCATAACTTGGTGAATTTGGGGCT 2385

QY      708  aAsnTyrArgArgAsnThrArgTyr-----GlyGluTrpAsnTrpTyrValLysAlaAs 726
Db      2386  TAACTATCAGCAACCAAGCCATCAAGCAGGCTCGGTTTCAGGTATTTTAAATGCTAA 2445

QY      726  pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db      2446  CAATCTATTAAACGATAAAGTCTTGTCTCATGAGACATTTTCCAGACATGCCACAAAT 2505

QY      746  tGlyArgSerPheThrGlyValAsnValLysPhe 758
Db      2506  GGGCGAAACTTTATGCTCGGGCAAAATTTCAAATTT 2542

RESULT 10
LOCUS   AX067459                      89047 bp      DNA      linear      PAT 24-JAN-2001
DEFINITION   Sequence 34 from Patent WO0078968.
ACCESSION   AX067459
VERSION     AX067459.1   GI:12545079
KEYWORDS
SOURCE      Moraxella catarrhalis
            Moraxella catarrhalis
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Moraxellaceae; Moraxella.
ORGANISM    1. 89047
            /organism="Moraxella catarrhalis"
            /mol_type="genomic DNA"
            /db_xref="taxon:480"
            66874
            /note="a or g or c or t, unknown, or other"
BASE COUNT  26501 a 17338 c 19060 g 26147 t
ORIGIN

```


Alignment Scores:			
Pred. No.:	5.09e-85	Length:	89047
Score:	1585.00	Matches:	338
Percent Similarity:	58.92%	Conservative:	141
Best Local Similarity:	41.57%	Mismatches:	272
Query Match:	39.27%	Indels:	63
DB:	6	Gaps:	13
US-09-936-377-2 (1-758) x AX067459 (1-89047)			
Qy	7	LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis	25
Db	424	AAACCGCTTGCTGTGCCATATTGGCAACTTTTCAATGCCAATGCTGGCAGAGCGCAAT	483
Qy	26	---GluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro	44
Db	484	TTAAAGGATGATGCAACCGCTCATTTTAGATGGCGTTTCGATCACTCTTTAGCTGACAA	543
Qy	45	ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp	64
Db	544	AATACAGATTGGCGTTAATCATTCATAAACAAGTCAGTGCGATCAAGTTTCAAAAGAG	603
Qy	65	ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis	84
Db	604	CAACTACAAACAACGAGCAACCACTAGGCGATGCCITGGCAGGTGAGCTGGCGTTCAT	663
Qy	85	AlaSerGlnIleThrGlyGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg	104
Db	664	TCTAACCATTTTGGGGCGGTGCTCAGCCCCCATCATTCGTGGCGAGGAGGTAAACGC	723
Qy	105	IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis	124
Db	724	CTGAAATCCTACAAACCGGTTACAGAGTTCTGGACATGTCGGTGTGTCGCACGACAT	783
Qy	125	AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr	144
Db	784	GCCATAGCGGTGGACACCACTCGCAAAACAGGTGGAGATTGTGCGAGGCTCTGTGGCC	843
Qy	145	LeuLeuIleThrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro	164
Db	844	TTGTGTACGCCTCTGGCACTCAGCAGGCGGTGTCATTCGTGTGATGACAAATACCC	903
Qy	165	GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn	184
Db	904	AGCAAAATGGCCAGCAAAA---TTACAAGGTGATGTGACGGTGGCGTCTTAGCAGTGCCAAC	960
Qy	185	LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnIleValLeuHis	204
Db	961	CGTGAAAAATTAATCACCGCGAGTGGCCGAAAGCCCACTGGGAGAGCATGTGGCAGTCCGT	1020
Qy	205	ThrGluGlyLeuIleThrArgLysSerGlyAspIleAlaValProArgIleThr---	220
Db	1021	GTTCAGGGGTGTCNAACAGCAGCAGACACTATAAAGCCGACCGCTTTGACCGCCATGTC	1080
Qy	221	-----	221
Db	1081	TTTAAACAAAAACATGAAGATGATAACACTCAGCGAGATTTCATCTATAAAGACACCTTA	1140
Qy	224	LysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGly	243
Db	1141	AAGCATCTGCCAGACACCATGCCAAATCAAAACCGAGACAAATATGCGCTGCC	1199
Qy	244	TrpArgLysArg-PheIleThrArgThrIleThrSerAspArgAspGlnIleThrGlyLeuPr	263
Db	1200	TGGCAATCAAGGCTTTTGGGGGCATCGGTGAGCTTACGCCGAGACAAATATGCGCTGCC	1259
Qy	263	AlaHisSerHisGluIleThrAspAspCysHisAlaAspIleIleIleThrGlnLysSerLeuI	283
Db	1260	CAACCATTCATGAATATGAAGAAATATAGCGTGCATGGCATTTCTCAGTCGCGCTTACA	1319
Qy	283	eAsnLysArgIleLeuGlnLeuIleThrProHisLeuLeuThrGluGluAspValAspThrAs	303

Db	1320	ATACAAGCCATATTGGCGTTTGGTATCTCTTTTGTATGAGAAATATGATGACTTACAGTTTGA	1379
Qy	303	pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi	318
Db	1380	CAATGGAGGGCTTGAATGCCATACATGATGACCCAGCACGACGACCATGCTCA	1439
Qy	318	sAlaHiSaAlaHis-----AsnGlyLysProTyrPheLeuArgAsnLys	333
Db	1440	TGACACAGCAGCAGCACGACGACGACCATGCGCAAACTTGGATTGATTTGAAATGAA	1499
Qy	333	sArgTyrGluLeuArgAlaGluThrLysGlnProPheProGlyPheGluAlaLeuArgVa	353
Db	1500	CGGTTATGATGTGCAAGGCAAAATCAATGCGCGGTTTGTGGCATTGATAAAATCCGAGC	1559
Qy	353	lHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPh	373
Db	1560	CAGCATGGGCAAAAGTCGATTATCATCATGATGAGATAGATGGGGGTGAGNAGCAGCTT	1619
Qy	373	ePheAsnAsnGlnThrGlnAsnAlaArgileGluLeuArgHisGlnProIle	390
Db	1620	TTTTTGATATCAAGCCAAATGTGTGGGCTCTGGAAGCTCACATACCCCATTCATACGCC	1679
Qy	391	---GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe	409
Db	1680	GATGGGCAAGTTTAGCGGGGTGTGGGGTAGGTTATCTCACCTCAAAAACACGCGACT	1739
Qy	409	u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs	424
Db	1740	TGTGCCACCTCGTTATCAAGATGGCAATAAACAAGACACCCAAACATCTTGCACAAATA	1799
Qy	424	nLysValGlnHisTyrSerPheGlyValGluGlnAlaAsn---TrpAspAsnPheth	443
Db	1800	TAAACCAAAACAGCGCAGTGTGTTTGGTTTGAAGAATACAAACCCCAATGACAAGCTGAC	1859
Qy	443	rLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe	463
Db	1860	CGTTGACGCGCGCTCGCATTCAGAAACAAACCATCACCTGGATTATGATAAAGACGC	1919
Qy	463	uile-----AspArgGluAsnTyrTyrLys	471
Db	1920	CATTATCAGAGCTTAACCTTAGCTTAGCAACCGCTCATGAACCAAGACATACGCTTTAA	1979
Qy	471	sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe	489
Db	1980	ACGATTGCTGGACAGCGGTACTTTAAACCCCAAAACAAACCGCACGCTCTTATGCTGT	2039
Qy	489	uSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl	509
Db	2040	TGGACGCATTTACAATTAACGCCCAACATAAATTTATCGCTGAATCTGTCGCATCAAGA	2099
Qy	509	uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh	529
Db	2100	ACGCTTGCCAAATGCTCAGGAATTGTATGCTCAGGCATGCACTTGCACCAACTCGTT	2159
Qy	529	eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl	549
Db	2160	TGAATTTGGCAACCGCTTTTAAACAAAGAAAAATCCACACATGATTGGGGCTGAC	2219
Qy	549	yTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy	569
Db	2220	ATTTCAAGGTGATAAATGGGATTATCGCTTGGGGCTATCATTTATGATTTTGATAACTA	2279
Qy	569	rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe	588
Db	2280	TGTGTTTTTACAACCAATTGTCGCGATGATAAGCAAGGTTTGGCGTCATCCGTCATGATA	2339
Qy	588	rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyValaGluGlyGl	608
Db	2340	AGACTTTAAAAACCGCACGCTATCAACACGACGGCGAAATTTTATGGATTGTATGTCA	2399
Qy	608	uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl	628
Db	2400	CATCGGTTATCAGATTAAATGATGATATCATGTGGGCTTATTTGGTATATATTCCTGG	2459


```

Db      697  TGTCCGTGGCGACCTAAATCCCGTTGTTACGTTTCATTAAAGCGATGATTATCCATTACGC 638
Qy      592  LeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyLuleTyrPhe 611
Db      637  TTATATCCGCTACATCAGCGAAGCAAGATTTATGGTTGGAAGCGAGTGCAT 578
Qy      612  LysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyLeuLys 631
Db      577  CAATTTCTCTCAACACAGAGTGTGCCATTTTGGTGATTATGTCGGGGTAAGTTAAAA 518
Qy      632  AsnLeuProSerLeuPro-----GlyArgGluAspAlaTyr 643
Db      517  GACCTTCGCTCTCTCCCGATTGGATATAAATATATCTATAACGAAATTTATGACATGTT 458
Qy      644  GlyAsnArgPro-----PheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
Db      457  GGTGTACACCAACCAACGGTTGGGAAAAACAACCTGATGTAATGCACCTCGTATGCACCA 398
Qy      662  AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
Db      397  ATCGCTTTTAGGCATAAAAATGGAATCGTATTTTGACAAACGGCATAAAGTTTCAATACGCAG 338
Qy      682  TyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
Db      337  CTTTATCGAGCTTTTGCAAAACAAAGTGGCCACCTAGAACCAACCAACGAAGGGCAT 278
Qy      702  HisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLuleTyrAsnTrp 721
Db      277  ACCATGTTGAATCTTGGTATGAGTTATGATGTAAGATGGCAATAATGATGATACCTTA 218
Qy      722  TyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSer 741
Db      217  TTTGCTAATGTAATAATGTGCTAAATAGTAGAGTCTATAATCATCTCGTTCTTATCC 158
Qy      742  AspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      157  TATATCCGCAAGTGGTCTAGGCTTAATGTTGGAATGAATTTAAATTC 107

RESULT 12
AE006179/c 11375 bp DNA linear BCT 08-MAR-2001
LOCUS      Pasteurella multocida PM70 section 146 of 204 of the complete
DEFINITION genome.
ACCESSION AE006179 AE004439
VERSION   AE006179.1 GI:12721801
KEYWORDS  .
SOURCE    Pasteurella multocida
ORGANISM  Pasteurella multocida
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Pasteurella.
1 (bases 1 to 11375)
May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
Kapur,V.
Complete genomic sequence of Pasteurella multocida, PM70
Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
21145866
11248100
2 (bases 1 to 11375)
Zhang,Q. and Kapur,V.
Direct Submission
Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
55108, USA
FEATURES
    source      Location/Qualifiers
1..11375
    /organism="Pasteurella multocida"
    /mol_type="genomic DNA"
    /strain="PM70"
    /db_xref="taxon:747"
157..465
    /gene="cyay"
    /note="synonym: PM1425"

gene

```

CDS

157..465

/gene="cyay"

/codon_start=1

/transl_table=11

/product="Cyay"

/protein_id="AAK03509.1"

/db_xref="GI:12721802"

/translation="MNVTEPHQNTIEQWAKIEEQLEAQCDCVDCDTCGVSFISITPQDR
SQVWINKPELLELWASRVGFFHRYQNNWNTSNDGKDFWACLEACAAGHEI VHS
"

gene

431..1351

/gene="PM1426"

431..1351

/gene="PM1426"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK03510.1"

/db_xref="GI:12721803"

/translation="MPLMVKLFTLVKEVIRMYKAGINLYKKLILAVLGVFVSHGTQ
ANESDARPYKSAEVLTKSDAFIGLLGYBENYLMGTYSNRHFLKREKTKQDEIKFKI
SLALPLARGILLGNNSVLAASVTKSMFQLSNVDSSPPRETNYPQLFLANKTQYSLP
PGWTLDQVEGINSQNGRDDAEKLSWNKLYVRASAKONWTVETIKPMWRIRPEKAK
NDNPDITKRGHEDFVALGYIYHDHDFKLSHYNPI SNKGLEASYSYPIITKNIRFYT
QYNYGESLIDYQORIQRIGIGISLNNVP"

gene

1341..3239

/gene="recQ"

/note="synonym: PM1427"

1341..3239

/gene="recQ"

/codon_start=1

/transl_table=11

/product="recQ"

/protein_id="AAK03511.1"

/db_xref="GI:12721804"

/translation="MFSSLLSKDNKSGAVDPHEIPLKOTALDVLHVAHVGYQSFRRGQE
EVIDATLMGKDSLVIMATNGKSLCYQIPALCFEGTLTIVISPLISLMKQDVQDLIANG
IEADYLSNSQTTFEQQVQNKMSGLTKLYVSPKVMWTTTFHLSHCKVSVFAIDE
AHCISQWHDPRPEYTOGLGKSCFPHAPIMALTADHATROQILRHNLQSPHYI
GSFDRPNIRYTLVEKFKPMOLCHFLVLOGKSGI IYCNSSKVERIAESLRNKGVA
QAHAGLTSOREQVQAFQRDNVQVVAITAFMGINKSNRFRVVFHDLPRSESYY
QETGRARDDLPRAEAVLFYFPADYAWLHKILLEKPESPQRIEALKQAIGFAESQT
CRRLVLNLYFGEHQHQKQNCIDICLPPQYDGLIDAKVMSTTIYRIGQRFVHYIA
VRLGSLNKIKDNQHEQLSVYIGKSKSKEHWQSVIRQLIHGFIKOVDFHFNATLQ
TENAKPILRGEQPLSLAMPRISSLTSVAPORYAIAOYDKLFLRLKQIADKEN
IPAVIVENDATLQEMAOYQPTTKAEMLAINGVATKFERFAQPFQMLIQKHKVLVTH
EPPLSLSS"

gene

complement(3338..5755)

/gene="PM1428"

complement(3338..5755)

/gene="PM1428"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK03512.1"

/db_xref="GI:12721805"

/translation="MISRGCKVNRKFFAVLMMCCIPQVWANTEKKQIVFLDEISVESK
GAAPRSDPLSGLPKONDILVSKOKLTGSSTLGNALAGELSVHSNORGGSSAPVVRG
QEGVRLKILONGSDVIDMSQSPDHAIQVDTLLAEQVEIVRGASTLLYANASPAVIRG
VVKRIFTQLPKGYEVDNTRINTNSHEKLVTAALTFTGLGKHLALRVEELRGSNNY
HVFAFKUDKTLNTPVDQNTKNSGNYGVAFIGERYGVGFAYNFKRYGLPGRHKLND
SCAAHITGWNVRNDYILGYLPHLMDTDLVNTFHCGNSHMDGKSHNPYGHHDH
SIAGPLIDSYAKRYDIAEVKQPMKATEKISYSETRYKDEKGNIAVNLKPNNGY
NLAVEIPTHITAGLSGVIAQYOTOTSSANI PRIAPCSNNASDPCHKKQDPKSLTKY
GDRKSWALIENTOSQSPFAIEQLRWQDFLEIGVTEKQRIDIEYDRAMFLVRKRL
EGCDPSFFSPSGCRGSGTAPDPASTHDIRATSYGAI SNWMTPTDTLSLTSYSHNR
HPTPMELYHGHKLATVSPFHGNNRLKKEVSDNVEGLAYIGDKLSYKAVNYNDPKN
RIFQNTLNKSNLARNQSKAKYQVEGRIDYALTPELHMLGFGYVARGKLDLPP
TYRVDHVANSLEPVTQDAPRPVPMRLGLFRVNMWMTESLTSLSLEYTVYQOKKVP
LENQTAAYSLNIGVDYSRQIAGVNYQLFQANNVLRNKYTSHTSFLFPFQMGKRVIT
LGLNIHF"

gene

complement(5967..6779)

/gene="PM1429"

Db 5011 ATCGGTGACGAGGTATGTCGGTTCCTACAAATCTTCGTGCGAGAAATATGGGTGA 4952
 QY 263 ProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeu 282
 Db 4951 CCGGACATACCATAGTAGTAGTAGTGTGTCGCAT---ATTGGGGTGCAATGTA 4895
 QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
 Db 4894 CGTAATGACATATTATTATGAGGCTTTATCTCATTAATG-----CATGATCGGACTA 4841
 QY 303 AspAsnProGlyLeuSerCysGlyPhe---HisAspAspAspAlaHisAla 321
 Db 4840 GTAAATACGATTTCCATTGTGATCAATCATGATGATGAGGAAACACATCATAAAT 4781
 QY 322 HisAsn-----GlyLysProTrpIleAspLeuArgAsn 332
 Db 4780 CATCTTACCGTCACGATCATGACCATTCATTGCGCGTCCATTGATGATTTATGCT 4721
 QY 333 LysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArg 352
 Db 4720 AAACGTTATGACATACATGCTGTAAGTCAAGCAGCCGATGAAAGCGATTGAGAAATCAAA 4661
 QY 353 ValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsn 372
 Db 4660 CTCAGTTATTCGGAACCGTTATAACACGATGAAAGATGCGAATATCGCCGTGAAT 4601
 QY 373 PhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArg 392
 Db 4600 TTATTTAAAAAATAACGATATACCTCGCGTAGAGATTTTCCATACGCCCATGACGGGG 4541
 QY 393 LeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAla-----408
 Db 4540 TTGAGTGGCGTTATAGGGCGCGAGTATCAACACAAACGACGATCGGAATATTCGCGGT 4481
 QY 409 LeuSerAlaThrSerGluAlaValLysGlnPro-----MetLeuLeuAspAsnLysValGlnHis 428
 Db 4480 ATTGCACCATGCTCAATAATGCAAGCGATCTTGTTCATAAGAAAGCAACGATGATCGG 4421
 QY 420 TCCAAATCACTAAAGTGATCGTAAGTCATGGCCATTGATTGAAATACTCAATCACAA 4361
 QY 429 TyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyVal 448
 Db 4360 ATGAGTTTTTTGCTATGAAACAATTCGCTGCGCAAGATTTTTTATTGAAATTGGTGA 4301
 QY 449 ArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle-----464
 Db 4300 CGTACTGAAAAACACGCGATTGATTTGAATATGATCGCGCTTGGCTGTTAAAGTAAAG 4241
 QY 465 -----AspArgGluAsnTyrTyr-----470
 Db 4240 CGAAAGCTGGAAGGCTGATCTCAATATCGTCTTTTATAGCCCATCAGGATGCTGTCAA 4181
 QY 471 ---LysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
 Db 4180 GCGAGTTATCCACCACTGATTTGCTCTTATGATGCGCGCGACCTCTTATTTGCTGT 4121
 QY 490 SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
 Db 4120 GCTATTAGTTGGAATATGACGCTGACTACACCTTTCTTGACTTACTCACATAACGAA 4061
 QY 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPhe 529
 Db 4060 CGTCACCCCAACCGGATGTAATTTGTTATACCGGCAACATTTAGCAGCGGTGTCATTT 4001
 QY 530 GluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
 Db 4000 GAACATGGCAATCGTAATCTGAAAAAGAAAGTTTCGATAACTGGGAGTGGCTCTGCG 3941
 QY 550 TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyr-ArgAsnArgPheGlyAsnTyr 569

Db 3940 TATCTTGGTGCACAGCTAAGTTTATAAAGTGAATGTATATTACAAATGATTTTAAAAATCGG 3881
 QY 570 IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu 589
 Db 3880 ATTTTAAATCAACATTGAAC-----AAATCCGGTAAT 3848
 QY 590 MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIle 609
 Db 3847 TTATCTTTGAATCGTTATTAATCAATCAAAAGCGAAATATTATGCTGGAAGGCGGTATT 3788
 QY 610 TyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg 629
 Db 3787 GACTATGCGTTGACACTGAACCTGCACATGGGACTTTTGGTGATATTATGTCGTGGAATA 3728
 QY 630 LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
 Db 3727 CTGTATGATTTACCGCAACGCTACCGTGTGATCATGTGCGGAATAGCTTAGAGCCGTT 3668
 QY 650 AlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLys 669
 Db 3667 ACTCAGCCTGATCAAGACGACCCAGCTGTCGCGCAATGCGTTTAGCTTTTCGTGTGAAT 3608
 QY 670 AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsn 689
 Db 3607 ATGGAGATGACTGAGAGTTTAAACAGTTCACTCGAATACACTTACCTTTATCAACAAAG 3548
 QY 690 LysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsn 709
 Db 3547 AAGTAGCCCGTTAGAAAATCAACGCTGCTATAGTTATTATTAATATCGAGTGGAT 3488
 QY 710 TyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeu 729
 Db 3487 TATTCACGCAAAATAGCTGGGTAAATATTCAATTATTGTTCAAGCAATAATGTGTTA 3428
 QY 730 AsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSer 749
 Db 3427 AACCGTAAAGTTTATCTCATCTCTTTTACCCTTTGTACCGCAGATGGGCGGTAAAT 3368
 QY 750 PheThrGlyGlyValAsnValLysPhe 758
 Db 3367 GTGACTTTAGGATTAAACATCCATTTC 3341

RESULT 13
 AE006112/c
 LOCUS
 DEFINITION
 PASTEURILLA MULTOCIDA PW70 section 79 of 204 of the complete genome.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 PASTEURILLA MULTOCIDA
 PASTEURILLA MULTOCIDA
 BACTERIA; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 Pasteurellaceae; Pasteurella.
 1 (bases 1 to 10166)
 May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and Kapur,V.
 Complete genomic sequence of Pasteurella multocida, Pm70
 Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
 21145866
 11248100
 2 (bases 1 to 10166)
 Zhang,Q. and Kapur,V.
 Direct Submission
 Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
 University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
 55108, USA
 Location/Qualifiers
 1. .10166
 /organism="Pasteurella multocida"
 /mol_type="genomic DNA"
 /strain="PW70"
 /db_xref="taxon:747"

```

gene      complement (75.. .1130)
/ gene="modC"
/ note="synonym: PM0742"
CDS      complement (75.. .1130)
/ gene="modC"
/ codon_start=1
/ trans1_table=11
/ product="ModC"
/ protein_id="AAK02826.1"
/ db_xref="GI:12721039"
/ translation="MLKINVKKQGLQALBANLOIPARGVTALFGLSGSKTSILNIV
SGLVHPDEGYSINERVLVQSKAVCPAYORHIGYVQDARLPFPHVTVKGNLCYGIK
KIDAKLPDIYSLIGIGIHLKRYPTITSGGKQKRAVTRALLTQPEILLMDPFLSALD
LPKRELSYLETISKIDIPILYVTHSIBELLARAEYVLLDEGKVPADRLESWE
NPLFLPKLEDEQAVLSLPIPLHNTGTQVTLGLQOQQQWIKRAQQAEGVENIRICIK
GSDVLSLTQPEKTSIRNLLHGKYRIVERNRVDVQIEIDKPIMATISKWALEDLA
LQLGQVPAQKANSVL"
/ complement (1117.. .1845)
/ gene="modB_1"
/ note="synonym: PM0743"
CDS      complement (1117.. .1845)
/ gene="modB_1"
/ codon_start=1
/ trans1_table=11
/ product="ModB"
/ protein_id="AAK02827.1"
/ db_xref="GI:12721040"
/ translation="MFTEFLSLGTTTIOAVKLSLSVSLNAILWSLPVAILMAWLL
ARKNFYKSLVSLGILHPLVLPVGVIGVLLVAMRNGFPGKLYLYSFGSFAFSWNG
AVLAAVAVAPLVPVRAJRLSLESIDMKLEQAOTLGASAEVFTITLPLSLPGLAG
VVLGPARSLGEGATIAVSNIAGETQTIPLAMTFTIQTPGAEEQARLCVLAILLSL
ISLISEWLAKMKQKQCSHAKD"
/ complement (2077.. .2841)
/ gene="modA"
/ note="synonym: PM0744"
CDS      complement (2077.. .2841)
/ gene="modA"
/ codon_start=1
/ trans1_table=11
/ product="ModA"
/ protein_id="AAK02828.1"
/ db_xref="GI:12721041"
/ translation="MLKMLTVSLATLAFSLSAQAQVTVFAAASMTNALNIOAED
YKVKPDDELVPSPASSSTLAKQIEEGAPVDIFVSANTKMKHSDKGLTKDTEKVL
VGNELVLTAPASQLKEVDVAKGEMVAELKDSFLSCGDPDHVAGQVAKAELTKLW
DKVEAKLARGQDVRAALALVERAERAPLGIYSTDGKSKGVKAIQVGPADYKAVEIP
VALLKRDNAETREFNLDESAAKKVLVSYGFSVK"
/ complement (3009.. .5786)
/ gene="PM0745"
/ note="synonym: PM0745"
CDS      complement (3009.. .5786)
/ gene="PM0745"
/ codon_start=1
/ trans1_table=11
/ product="unknown"
/ protein_id="AAK02829.1"
/ db_xref="GI:12721042"
/ translation="WNVILFLYKKWIFLQNNIFMLNLSLTFVFKVVDVETKVV
TNKLIKQITLLPWSAVAFTHVIYAQTEMLDEIVVSGAOPHLGASAIHYQAIQNOVI
KBERLQKQATLGNALGELGVHSPFGGASAPILRQEGVRIKILQNGLDVMSA
TIGDHAVADSLLAQVGLRSTHLYSSAAGVNVNVDKRIPTAVPEKGVGEIIF
TRFDVQBSTAGITFRLHPLALRLGLKRYTHYRVFAFKSGGGETIRYLPDSHN
RSQVGTIGVSTWTKQSYLGVSYSEERDRYGLPGHNHRYDRCKSHVYDAAARPELGKY
LTPYPLAADTDPVFAHLDCGTCGIDNDPSSHDDPFGHEHDHSHGGPWRHLHGRFD
LRGOWDSPAMLDKVKGSFYADITHYEHYSGQAGTDKDRDSFTIERKKAEKNRGK
AAGIYKNSGVNGLRFEYHPTPIVGLSGVGVQVSEYKTAIAPLGSQIKHOHLVPTQ
KQASFFAVENYVDDPIFEVGARVDKQRIPIKYDQHVLAHAKKGGDHPDLSHKEKA
VSYLGVSWLFPNRYRIGLTSRNERLPTPMELIYHGOHLATNSFEHGNKNLRKESN
NIELGPAYTDKWDKLYSLYQNKFRYIYNEDLARYGNALFRRYTQARAKFHGEAEL
NFRPTDYQVTLFGVYGRFLPDLPEQYQFYQIYADEEGLIQANWEKQPYRIEG
IKRNEADVPAPRLGMRLSNVTEHLSFADYTVFVSQQTASSLSIKPRLALEAS
DFDIDDGTGNLLKLGIDRDKYNRGTQPIADLSADKADPESALEAIRLEENKLPAX
IEKIQDPSKSHLLNIGVNYQRIQIHLDSYGVFSVNNVLRQVTVHYSYLPYPMQMG
RNIVLNFVGK"

```

```

gene      complement (6825.. .6825)
/ gene="modE"
/ note="synonym: PM0746"
CDS      complement (6825.. .6825)
/ gene="modE"
/ codon_start=1
/ trans1_table=11
/ product="ModE"
/ protein_id="AAK02830.1"
/ db_xref="GI:12721043"
/ translation="MTLSEILLTIKLOQLPVDPKRIRLLKEIKQCGSINOAKNAK
VSYKANDHLAMNEISPKPFLERNVGNKSGSTLYVYAOPLLOLYDLLEHTQOKAF
HILQENPLDSVLHATKAFSLQSSARQFQDGI VALRHENIHCVATQIKGLAFLIT
VATIKSQRLKTLTGKVMKIKAPVAVKREKPEGVNTFLVNNKIEITDKGDVEEVI
LTSINKHDEMEFCATLNKAENLPDEQIWIISIDPEQVILATLY"
/ complement (6868.. .7455)
/ gene="PM0747"
/ note="synonym: PM0747"
CDS      complement (6868.. .7455)
/ gene="PM0747"
/ codon_start=1
/ trans1_table=11
/ product="unknown"
/ protein_id="AAK02831.1"
/ db_xref="GI:12721044"
/ translation="MKISHFFACLDRLIQRWSLMRNLEKENLAHSHQVAFVAHAL
AVIKNFYQCHLNPRIAMVAMYHDTSEIFTGDLTPPIKYFNPKITQAKYENRAEA
HLIALPLKELQADFPYLDSTQPSAQERHVVQKADLI CAIVKAQFELENGNQEFKAAK
ARLETLMKTHSDMDRYFTEVPFSGFKPIDEITL"
/ complement (7487.. .8515)
/ gene="degS"
/ note="synonym: PM0748"
CDS      complement (7487.. .8515)
/ gene="degS"
/ codon_start=1
/ trans1_table=11
/ product="DegS"
/ protein_id="AAK02832.1"
/ db_xref="GI:12721045"
/ translation="MIKKLPHSVMIGLVCAAVILVPSHVNQFSLKAPSPMSFKDAVR
VASPAVVNVNRSFSAEQLVGNLGSVIMSKDGYILTNKHVIONADQIVVALQON
QGLDASLTGSDTLTDLAVLKIRADNLSTIPQNPQRMHVGDVALAIGNPYNIGSVS
QGIISAIGNVAGDSVGRNQFTQDASINRNGSGGALINSVGLIGINTLSIGKTAGE
IAGELNPAIPELANDVMKRIIRDGRVIRGVQVSDILFNGGQHALEKGLITNVN
PEPGACKGAKGAGDILKVGVDVAISPAQMQVVISNTRPNTRLNVVVISRLKTIIELPV
IIEEYPLN"
/ complement (8512.. .9642)
/ gene="ribD"
/ note="synonym: PM0749"
CDS      complement (8512.. .9642)
/ gene="ribD"
/ codon_start=1
/ trans1_table=11
/ product="RibD"
/ protein_id="AAK02833.1"
/ db_xref="GI:12721046"
/ translation="NMTFTAQDVVVMQMALEAEKQQTTPNPFSVGCVLVKDGIIVG
RGFVRAGEAHEAVMLRAGENARGATVITLPCSHFQRTPPCAQGLIDAGVKKV
SLCDRNPQVAGRGLSMLDAGIESAVGLLAKEAELNRLGRLKMTGRYVQLKWMAM
SMGDRPMANGESKWTGELARADYQAEKASAILSTSATVLADPLLNVRWQOLPD
KVRADYTAQLRQFVRVILDSQHRVQPSHRLFSLSPVWLVSQTASRLLEFPFHCEQI
QIASANTADLMALIELGQRIQINTLWIEAGATLAGALIEAKLVDELIIYIAPKLGDRA
RGICKPLHSLTAAESLWQLHSTKLGDDLLKTIYHPRIETPT"
/ complement (9642.. .10091)
/ gene="PM0750"
/ note="synonym: PM0750"
CDS      complement (9642.. .10091)
/ gene="PM0750"
/ codon_start=1
/ trans1_table=11
/ product="unknown"
/ protein_id="AAK02834.1"
/ db_xref="GI:12721047"
/ translation="MECPFCSTETKVIDSLSEGVYVRRRECCNGHERFTTPTETA
ELIIPKVKNDGTREFNFKRGKIQHALEKRPVSENQVRAISYIIHRLRSTGBRE

```


Qy	29	GlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSer	48
Db	5579	CAAAACGGAAATGCTGTGATGAATGTAGTCAGTGGTCACAAACGGCACTTAGCTGGTTCA	5520
Qy	49	GlyLeuLeuHisThrSerThrAlaSerAspLysIleSerGlyAspThrLeuArgGln	68
Db	5519	CGCATAGAGCAATTCAAGCGATTCCCATCAGGTGATTAATAAAGAACGGCTGCAAAA	5460
Qy	69	LysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyr	88
Db	5459	CAATCTGCCACTTAGTAAATGCTTAGCAGGTCAATTAGTGTGCACTTCCAAACCCCTTT	5400
Qy	89	GlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeu	108
Db	5399	GGTGGCGGGCARGTGGCCAAATATCCGAGGACAGAAGGTGACGTATTAATAAATTTG	5340
Qy	109	AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetVal	128
Db	5339	CAAAATGGCTTAGATGTCGTGGATATGTCACGGATTTTCGCGGATACAGCTGTCGCGCC	5280
Qy	129	AspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSer	148
Db	5279	GATAGTTTATTAGCTGACAAAGTAGAATTAGTAGTGGTGCTCCACTTTGTTATATCT	5220
Qy	149	SerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetPro	168
Db	5219	AGTGCTTCTGCGCGGTGTAGTGAATGTCGTGGACAAACGTATTCACACAGCTGTACCT	5160
Qy	169	GluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeu	188
Db	5159	CAAAAGGGCTATGAAGGTGAATTTTACCCGTTTTCATACAGCAAGTCAGCAAAAGTACT	5100
Qy	189	ThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu	208
Db	5099	GGAACCGCAGGATCACITTTAGATTACATCTCATCTGCCTTACGTTTGGAAAGGCTTA	5040
Qy	209	TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsn-----LeuLys	224
Db	5039	AAAAGTTATTCCGACTCATATTATCGTGTACACGCTTTCAAATCAGGTGCTGAAACGATTCGT	4980
Qy	225	ArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyr	244
Db	4979	TATTTGCCAGACAGTCATACCCCTCACAACTGGGA-----ACATTTGGTGTTCCTTGG	4926
Qy	245	ArgLys-----ArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeu	262
Db	4925	ATTAGGATCAATCTTATCTAGGGGTATCTTATAGTGAACGACGCGATCGCTATGTTTG	4866
Qy	263	ProAlaHisSerHisGluTyrAspAspCysHisAlaSplIleIleTyrGln-----Lys	280
Db	4865	CCTGCTCATATATCAATAATATGATCGCTGTAAGGCCCATGTGGTTGATGAAGCGGCAGA	4806
Qy	281	SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluAspVal	300
Db	4805	CCTGAGCTTGGTAAGGCTATCTGACACCTTATCCATTTAGCAGATGATACAGATATT	4746
Qy	301	AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
Db	4745	GTGTTTGCACACTTAGATGGATGTATGGTGTATGATATGATCTCTCCATGAGCAT	4686

QY 645 -----AsnArgProPhe----- 648
Db 3632 ATTGCTTATGATGAGGAGGAGCTGATTCAAGCAAAATGGGAAAAACACCTTACCGTATT 3573
QY 649 -----IleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPhe 666
Db 3572 GAAGGCATTAAACGTAATGAGCGTGATGCGCTCGTGTGCCACCGCAGCTTTAGGTATG 3513
QY 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPhe 686
Db 3512 CGCTGAGTGGCAATATAACGGAACATTTATCTTCTTCTTGGCAATATACATTATGCTCTTT 3453
QY 687 AlaGlnAsnLysLeuAla----- 692
Db 3452 TCTCAACAAATAACTGCTCTTCTTCTTCAATTAAACCGCGAGAGCGTTAGAACGAGT 3393
QY 693 -----Arg 693
Db 3392 GATTATTATGATGATGATACAGCGAAATTTATTGTTAAAGGTATTGACCGGACAAA 3333
QY 694 TY----- 694
Db 3332 TATAATCGTACGGGTGAGCTATCGCAGATTAAAGTGCTTCTGATATAAGCGGATTTGAA 3273
QY 694 ----- 694
Db 3272 AGTGGCTGGAAGGAGTTCGTTTAAGAGAGAGAAAAATGAAATAATTCCTGCTAAAAATAGAA 3213
QY 695 -----GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 3212 AAAATTCAGAGATCCGAGTAAGGCGCATCTTTGTTAAATATAGTGTAAATATCA 3153
QY 712 ArgAsnThrArgTyrGlyGluThrAsnTyrTyrValLysAlaAspAsnLeuLeuAsnGln 731
Db 3152 CGTCAAAATTTGGCATTTAGACTATTCTGTCGGTTTCAGTGTCTCAATAATGTTATTAACCAA 3093
QY 732 SerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThr 751
Db 3092 CGTGTATTATGTTACACCTCTTATTACCTTATGTCGCACAAATGGGGCGCAATTATG 3033
QY 752 GlyGlyValAsnValLysPhe 758
Db 3032 CTGAATTTTGGGGTAAATTC 3012
RESULT 14
AX394912
LOCUS
DEFINITION Sequence 49 from Patent WO0218595.
ACCESSION AX394912
VERSION AX394912.1 GI:21065985
KEYWORDS
SOURCE Moraxella catarrhalis
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
REFERENCE 1
AUTHORS Loosmore,S., Wang,J., Bradley,B., Ochs,M. and Yang,Y.P.
TITLE Moraxella polypeptides and corresponding dna fragments and uses
JOURNAL thereof
PATENT Patent: WO 0218595-A 49 07-MAR-2002;
FEATURES Aventis Pasteur Limited (CA)
source Location/Qualifiers
1. .2942
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
101. .2842
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD32082.1"
/db_xref="GI:21065986"
/translation="MIKKPLVCAISATFAMPVADNTKLGEETTTILKGVLVSSQTNQ"

NTGFVNSDKSSDLSKDKLKYRSATLGNALSGELGHSNPFQGGSSAPVVRGQG
VRLKILQNGTVIDVDSISPDHVVATDILLASKVELVRGADTLLYGLASPAGVINVD
DRPNMPSGAIHDKIEGTMLRYNNHKEKATAGVSGVGDRIAVRVEGLKREADD
YVPHFOADPMLDYVPGSANNSTVMIGVSYHDNGHIGASYSHRDXYGIPCHIBCD
SQREHLKWHNTKTSNYLPIYPLHEDSDIDNPHTHCHNHEDHIGENHPTGVPIN
HEHHSWIDMTKTRIDIRGEVYRPIQLODKIKLSLYADYHDEKADAGNEQDNNHKP
SERDITVDKGHASSIFTKGVNRLSYHTFKRLSGVLGIGYQTKQSRAMHRBKTP
FQSEAWQKAQSQNIQYRPLVNPNTKSLGIFGLEQLKLNQMTFKVAMHRBKTP
IEYDQHLHALQVFLSKAQLPDHPDLTYYKQHATSYAGSALWDITNHRLSLTY
HNRIIPSMELYQGGHLATSPHGNKLVKEKSDNYELGFMHTADKYSKASTYYS
NFNYLPHETIAKEGNYLPRVNTAKTVEGSLTYOPNANHSVFMCDVQMGKIG
ALSDYKGLVIAGRKVYEDDDIKMTVDNDGVDYDAGGLTCAKLTPEOMGQNDND
CSTIYNKKGTTTSGEEDYDLARNPTAPRVPFVPSRLGRWQGYGFDHWSANAENH
VFLQNKVATSTVAIKFQKQEGORHESCRISDYGSNNPLMMQPRYTENKTAGY
NLLNVGLDYNNAVYNDYTLIRANLLNEQIYHNSFLFPVPMQGRNLTGLTAKP"
BASE COUNT 911 a 562 c 632 g 837 t
ORIGIN
Alignment Scores:
Pred. No.: 6.44e-64 Length: 2942
Score: 1205.50 Matches: 290
Percent Similarity: 46.25% Conservative: 142
Best Local Similarity: 31.05% Mismatches: 303
Query Match: 29.87% Indels: 199
Gaps: 18
US-09-936-377-2 (1-758) x AX394912 (1-2942)
QY 9 IleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlu 28
Db 83 ATCATACCGAGGATCATCATGATAAAAAACCACTTGTGTCGATATCGGCCACCTTT 142
QY 29 GlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSer 48
Db 143 GCGATGCCAGCGGTAGCAGATTAATACCAAGCTGGGTGAAGAGCCAAACCCACCTTAAAG 202
QY 49 GlyLeuLeuHisThrSer----- 54
Db 203 GGTGATTGTGTAAGCTGCACAAACGAAACCAAAATACAGGTTTGTATCTAATGATTCAAA 262
QY 55 ThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGly 74
Db 263 CAATCCAGTGATCTTACGCTTTCAAAAGATAAATTAATAATATCGTCGGCAACCTTGGGC 322
QY 75 AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyValAsnAla 94
Db 323 AATGGTTAAGTGGTGGAGCTTGGTATTCATAGTAACCTTTTGGTGGCGGTCACTGCA 382
QY 95 ProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThr 114
Db 383 CCTGTTGTGCGAGGCAAGAGGTTGCGTCTTAAGATTTTACAAAATGGAACGTGATGTG 442
QY 115 GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGln 134
Db 443 ATGTGATGTGTCATCAATATCGCTGATCATGTTGTGGCGACCGATACATCTTTTACGCT 502
QY 135 GlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGly 154
Db 503 AAAGTTGAGCTTGTGCTGCTGCTGATACGCTGTTATATGCTTGGCATCCCGACCTGGT 562
QY 155 LeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly----- 171
Db 563 GTGATTAATGTTGTTGATGACCGTATCCGAATCGTATGCTGTGCTGCTATCCATGAC 622
QY 172 ---ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer 190
Db 623 AAAATCGAAGCGGACGATGCTTCGATATTAACACAAACACCATGAAAGCTTGCACACT 682
QY 191 GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg 210
Db 683 GCAGGGGTGAGCTTTGGGGTAGGAGATCGCATTTGGGTTCGGGTGGAGGGCTTAAAGCGA 742
QY 211 LysSerGlyAspTyrAlaValProArgTyr-----ArgAsnLeuLysArgLeuPro 227

```

Db      743  GAGCGTGATGACTATCAAGTTCCCATTTTTCAGGCGAGATCGCATGTGTAGATTATGGCCA 802
Qy      228  AspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGly----- 243
Db      803  GGTAGT-----GCAATAACTCTACCGTTGGCATGATTGGCGTGTCTTATATT 850
Qy      244  TrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro 263
Db      851  CATGATAATGGGCATATCGGTGCTTCTATAGCCACCGTAAGATCGTTATGGTATCCCA 910
Qy      264  AlaHisSerHisGlnTyrAspAspCysHisAlaAspIleTyrGlnLysSerLeuIle 283
Db      911  GGGCATATCCACTCGCGACGCCACGAGAGCATTTTATCAATGGCATAATATCAAAA 970
Qy      284  AsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAsp 303
Db      971  TCCAAATATTATTTACCATTATTCCTCATTTGATGAGGATTCAGATATTGAT--GAC 1027
Qy      304  AsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAsn 323
Db      1028  AATCCTCATAGCATTCGCGCCACACACGAGACCATATCGTGAGCATATCCACG 1087
Qy      324  Gly-----LysProTrpIleAspLeuArgAsnLysArgTyr 335
Db      1088  GCGCGTCCCATCAATCATGAACATCATTCGCTTGGATTGATATGAAAAACCAATCGCTAC 1147
Qy      336  GluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu 355
Db      1148  GACATTCTGGCGAGGTATATCGGCTTATTCAGGTTTGGATAAAATTAAGCTTAAGCTT 1207
Qy      356  AsnArgAsnAspTyrHisAspGluLys--AlaGlyAspAla----- 369
Db      1208  ACTTATGCAGATTATTATCATGATGAAAAAGATCTGGCAATGAGCAAGACCCAAAACAAT 1267
Qy      370  -----ValGluAsnPheAsn 375
Db      1268  CAAAAACCTTCTGAGCGTGATACAAACGCTGGATAGGTCATGCCAGCTCTATTTTTACA 1327
Qy      376  AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly 395
Db      1328  AAAAAAGCGTTAATGTGTGCTGGAGTTATATCATACCGACCAACGCTTATCTGGG 1387
Qy      396  SerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla----- 411
Db      1388  GTATTGGTATTTGGTATCAAAACCCAAATCTCGACGAGAGGCGTATTTGCCAAGC 1447
Qy      412  -----ThrSerGluAlaValLysGlnPro----- 419
Db      1448  TATTTTCAATCAGAAGCAGATGCGAAAAAGCCCAAAAGTCAAAACATTAAACCAATATCGT 1507
Qy      420  -----MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAla 437
Db      1508  CTTTACTATTATAGTTCCAAATACCAATAAAAGCCTTGGTATTTTGGACTGTGAGCACTA 1567
Qy      438  AsnTrpAspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIle 457
Db      1568  AAGCTAATCAATGACTTTTAAAGTGGCGATGCTCATGAACACAAAACACCAAT 1627
Qy      458  ArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr----- 470
Db      1628  GAATATGACAGCATTTACTTTGACCATGCTTTTGGCAGTATTTTAAAGTAAAGCACAGCTA 1687
Qy      471  LysGlnPro--LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
Db      1688  AAAGCACCTGATCATCTCTGATTGACGACATATAAACACATGCCACTTATATGCTGGT 1747
Qy      490  SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
Db      1748  AGTGCCTTATGGCATATTACGCCAAATCATCGATTGTCATTGACCTACTCACATAAGCAA 1807
Qy      510  ArgLeuProSerThrGlnLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPhe 529

```

RESULT 15

AX067466

LOCUS

```

Db      1808  CGCATTCATCGCGCATGAGCTGATTATATCAAGCGGACACATTTTGGCGACGACTCTTTT 1867
Qy      530  GluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
Db      1868  GAGCATGGCAATAAAACTTTGGTCAAAAGAAAATCGGATTAATATGAGCTGGGTTTTATG 1927
Qy      550  TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
Db      1928  CATACAGCAGATAAAGTCAGCTATAAAGCAAGCACTTACTATACCAATTTTGTAACTAT 1987
Qy      570  IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu 589
Db      1988  ATTTTAAATGAGACC-----ATTGCCAAAAGAGCAAT 2020
Qy      590  MetLysLeuValArgTyrAsnGlnSerClyAlaAspPheTyrGlyAlaGluGluIle 609
Db      2021  TTATACATCAGACGCTATAATACAGACGCGCTAAGTTTNGTGTGGAGGGTTCATTA 2080
Qy      610  TyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgLysArg 629
Db      2081  ACTTACCAGCAAAATGCCAATCACAGTGTGATGTTTTTGGTGATATGGTCAAGGTAAA 2140
Qy      630  LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db      2141  ATTGGTGCATTATCTGATATTAAAGGTAAAGCTT---GTATATGCTGGCAGAAAGTGGGT 2197
Qy      650  AlaGlnAspAspGln----- 654
Db      2198  TATTTTGTGATGATATCATCAGGATATGACCGTTGTGATGATATGTTATGATCCGAT 2257
Qy      654  ----- 654
Db      2258  GGTGTTTGTGACTTGTGCTTAAACCGCTGAAACAATGGGACAAAATTAACGATAATAAT 2317
Qy      654  ----- 654
Db      2318  GATTGTAGTACACCACTTAATGTCTATAAAAACGGTACACAACCTCAGGCGAGGAGAT 2377
Qy      655  -----AsnAlaProArgValProAlaAlaArgLeuGly 665
Db      2378  TACGACCGTTTGGCAGCAAAATCCAATTCAGCACCCAGAGTACCGCCCGCGTGGGC 2437
Qy      666  PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db      2438  ATTCGTTGGCAAGGATTTTGGTGTATCATTTGCTCTGCCAATGACAGAAATTAACCATGTG 2497
Qy      686  PheAlaGlnAsnLysLeuAla----- 692
Db      2498  TTTGCACAAAATAAAGTTTGGCCACCTCAACGGTTGGCATTTAAACCTCAATTCACGACCA 2557
Qy      692  ----- 692
Db      2558  GAAGGTTGCCAACGCCATGAGAGTCATTTGCCGAATACGCGACTATGGCAGTGTAAACAAC 2617
Qy      693  -----ArgTyr-----GluThrArgThrProGlyHisMetLeu 704
Db      2618  CTTTGTGATGATGACCAAGATATATCACAGAAACAAACCGCAGGATATAATTTGCTC 2677
Qy      705  AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluThrPAsnTrpTyrVallys 724
Db      2678  AATGTTGGCTAGATTATAACAATGATATGATGTTGATTATACGCTGTCAATTGCT 2737
Qy      725  AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro 744
Db      2738  GCGAATAATTTACTTAATGAACAAATCTACATTCACAACTCATTTTGGCGTTGTACCG 2797
Qy      745  GlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
Db      2798  CAGATGGGCGTAATCTGACTTGTAGTTTACGGCGTAAATTT 2839

```

269223 bp

DNA

linear

PAT 24-JAN-2001

DEFINITION Sequence 41 from Patent WO0078968.
 ACCESSION AX067466
 VERSION AX067466.1 GI:12545091
 KEYWORDS
 SOURCE Moraxella catarrhalis
 ORGANISM Moraxella catarrhalis
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Moraxella.

REFERENCE 1
 AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
 TITLE Nucleotide sequences of moraxella catarrhalis genome
 JOURNAL Patent: WO 0078968-A 41 28-DEC-2000;
 INCYTE Incyte Genomics, Inc. (US)

FEATURES
 source
 1..269223
 /organism="Moraxella catarrhalis"
 /mol_type="genomic DNA"
 /db_xref="taxon:480"

BASE COUNT 77067 a 56596 c 57380 g 78180 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.99e-61 Length: 269223
 Score: 1205.50 Matches: 290
 Percent Similarity: 46.25% Conservative: 142
 Best Local Similarity: 31.05% Mismatches: 303
 Query Match: 29.87% Indels: 199
 DB: 6 Gaps: 18

US-09-936-377-2 (1-758) x AX067466 (1-269223)

Qy	9	IleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlu	28
		: : :	
Db	209651	ATCATACCGAGGGTATCATGATAAAAAACACCTGTTGTGGCATATCGCCACCTTT	209710
Qy	29	GlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSer	48
		: : :	
Db	209711	GCATGCCAGCGGTAGCAGATAATACCAAGCTGGGTGAAGAGCCACCACTTAAAG	209770
Qy	49	GlyLeuLeuHisThrSer-----	54
		: : :	
Db	209771	GGTGTATTGGTAAGCTCGCAACGAACCAAAATACAGGTTTGTATCTAATGATTCAAA	209830
Qy	55	ThrLaserAspLysIleIleSerGlyAspThrLeuArgGlnIleAlaValAsnLeuGly	74
		: : :	
Db	209831	CAATCCAGTGATCTTACGCTTTCAAAGATAAATATAAATATCGTTCGCAACCTTGGC	209890
Qy	75	AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAla	94
		: : :	
Db	209891	AATGCTTAAAGTGTGACCTGGTATTATAGTAAACCCCTTGGTGGCGGTTCATCTGCA	209950
Qy	95	ProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGlyGluThr	114
		: : :	
Db	209951	CCTGTTGTGCGAGGCAAGAGGGTGTGGCTCTTAAGATCTTACAAATAGGAACCTGATG	210010
Qy	115	GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGln	134
		: : :	
Db	210011	ATTGATGTCATCAATATCGCTGATCATGTGTGGCGACCATACACTTTAGCGTCT	210070
Qy	135	GlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGly	154
		: : :	
Db	210071	AAAGTTGAGCTGTTTCGTGGTGTCTGATACGCTGTTTATATGGCTTCGCATCGCCAGCTG	210130
Qy	155	LeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly-----	171
		: : :	
Db	210131	GTGATTAATGTTGTGATACCGCTATCCGAATCGTATGCTTCCTAGTGGTGTATCCATGAC	210190
Qy	172	--ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer	190
		: : :	
Db	210191	AAATCGAAGGCGAGCGATGCTTCGATATAACACCAACCATGGAAGCTTGCACCT	210250
Qy	191	GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg	210
		: : :	

```

510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPhe 529
Db 211376 CGCATTCATCCCGGATGGAGCTGATTATATCAAGCGGACATTTGGCGACCGAGCTTTT 211435
QY 530 GluValGlyAsnLysHisLeuAsnLysGluArgSerAsnLysLeuAlaLeuGly 549
Db 211436 GAGCATGGCAATAAAACCTTGTCAAGAGAAATCGGATATATGAGCTGGTTTATG 211495
QY 550 TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
Db 211496 CATACAGCAGATAAAGTCAGCTATTAAGCAAGCCTTACTATAGCAATTTTGAATCAT 211555
QY 570 IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu 589
Db 211556 ATTTTAAAGAGCC-----ATTGCCAAGAGGCAAT 211588
QY 590 MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyLeu 609
Db 211589 TTATACATCAGACGCTATTAATCAGACGCGGTAAGTTTATGTTGGAGGGTTCAATTA 211648
QY 610 TyrPheLeuProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg 629
Db 211649 ACTTACCGCAAAATCCCAATCACAGTGTGATGTTTGGTGATATGTCGAAGGTAAA 211708
QY 630 LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db 211709 ATTGGTGCATTATCTGATATTAAGTAAGCTT---GTATATGCTGGCAGAAAGTGGGTT 211765
QY 650 AlaGlnAspAspGln----- 654
Db 211766 TATTTGATGATGATATCAAGATATGACCGTTGATGATATGATGATGATGATGATGATGAT 211825
QY 654 ----- 654
Db 211826 GGTGGTTTGACTGTGCTTAAACCGCTGAACCAATGCGGACAAATTAACGATAATAAT 211885
QY 654 ----- 654
Db 211886 GATTGTAGTACAAACCAATTAATGCTATATAAAACGGTACAAACCTCAGCGAGGAGGAT 211945
QY 655 -----AsnAlaProArgValProAlaAlaArgLeuGly 665
Db 211946 TACGACCGTTTGGCAGAAATCCCACTTACGACCCAGAGTACCGCCGCGCTGGGC 212005
QY 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db 212006 ATTCGTGGCAAGGGTATTTTGGTGATCATTTGCTGCAATGCAAGATTTAAACCATGTG 212065
QY 686 PheAlaGlnAsnLysLeuAla----- 692
Db 212066 TTTGCACAAATAAAGTTGCGCACTCAACGGTTGCCATTAAACCTCAATTCAAGCAGCCA 212125
QY 692 ----- 692
Db 212126 GAAGTTTGCCACGCCATGAGCTCATTCGGAATCAGCGACTATGGCAGTGATACAAAC 212185
QY 693 -----ArgTyr-----GluThrArgThrProGlyHisMetLeu 704
Db 212186 CCTTTGATGATGAGCAGCAATATATCAGAAACAAACAGCGCATATATTTGCTC 212245
QY 705 AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLys 724
Db 212246 AATGTTGGCTTAGATTATAACAATGATCATATCGTAATGTTGATATACGCTGCAATTCGT 212305
QY 725 AlaAspAsnLeuAsnGlnSerValTyrAlaHisSerPheLeuSerAspThrPro 744
Db 212306 GCGATATTTTACTTAAGTAAACAAATCTCATTTCAACATCAATTTTGGCTTTGTACCG 212365
QY 745 GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 212366 CAGATGGGGCTAATCTGACTTTAGGTTTGACGCGCTAAATTT 212407

```

RESULT 16
 AX052569
 LOCUS 2742 bp DNA linear PAT 12-JAN-2001
 DEFINITION Sequence 11 from Patent WO0071724.
 ACCESSION AX052569
 VERSION AX052569.1 GI:12226795
 KEYWORDS Moraxella catarrhalis
 SOURCE Moraxella catarrhalis
 ORGANISM Moraxella catarrhalis
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Moraxellaceae; Moraxella.
 REFERENCE 1
 AUTHORS Thomard,J.S.
 TITLE Novel compounds from moraxella catarrhalis
 JOURNAL Patent: WO 0071724-A 11 30-NOV-2000;
 SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
 FEATURES
 source
 1. 2742
 /organism="Moraxella catarrhalis"
 /mol_type="genomic DNA"
 /db_xref="taxon:480"
 BASE COUNT 844 a 536 c 600 g 762 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 9.59e-64 Length: 2742
 Score: 1202.00 Matches: 291
 Percent Similarity: 46.66% Conservative: 142
 Best Local Similarity: 31.36% Mismatches: 301
 Query Match: 29.78% Indels: 194
 DB: 6 Gaps: 20
 US-09-936-377-2 (1-758) x AX052569 (1-2742)
 QY 7 LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
 Db 10 AAACCACTGTTGTGGC-----ATATCGCCACCTTTCGATGCCAGCGGTAGCA 60
 QY 27 ThrGluGlnSerValGlyLeuGluThrValThr-----ValValGlyLysSer 42
 Db 61 GATAATACCAAGCTGGGTGAAGGCCAACCCACCCTTAAAGGGTGTGTTGGTAAGCTCG 120
 QY 43 ArgProArgAlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAspLysIle 60
 Db 121 CAACGACGACCAATACAGGTTTGTATCTAATGATTCAAAACAATCCAGTATCTACG 180
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 181 CTTTCAAAAGATAAAATTAATAATATCGTTCCGCAACCTTGGGCAATGCGTTAAGTGTGAG 240
 QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 241 CTTGGTATTCATGTAACCCCTTTGGTGGCGGTTCATCTGCACCTGTTCGCGAGGCA 300
 QY 101 ThrGlyArgGlyLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 301 GAGGTGTGCTCTTAAGATTTTACAAATGGAATCATGTGATGATGATGATGATGATGAT 360
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLeuArg 140
 Db 361 TCGGCTGATCATGTGTGGCGGCGGATACACTTTTAGCGCTTAAAGTTAGCTTTGTCGT 420
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 421 GGTCTGATGATGCTGTTATATATGCTTGGCTTGGCATCGGCTGCTGCTGCTGCTGCTGCT 480
 QY 161 GlyLysIleProGluLysMetProGluAsnGly-----ValSerGlyGluLeu 176
 Db 481 GACCGTATCCCGAATCGTATGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 540
 QY 177 GlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGly 196
 Db 541 ATGCTTGTGATATACAAACCAACCATGAAAGCTTGCACCTGCGAGGGGTGAGCTTTGGG 600

Db 2716 ACTTTAGCTTTCAGCGCTAAATTT 2739

RESULT 17
 AE012378
 LOCUS
 DEFINITION
 Xanthomonas campestris pv. campestris str. ATCC 33913, section 286 of 460 of the complete genome.
 ACCESSION
 AE012378 AF008922
 AE012378.1 GI:21113832
 KEYWORDS
 Xanthomonas campestris pv. campestris str. ATCC 33913
 SOURCE
 Xanthomonas campestris pv. campestris str. ATCC 33913
 ORGANISM
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xanthomonas.
 REFERENCE
 1 (bases 1 to 10947)
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.B.A., Camarotte, G., Cannavan, F., Cardozo, J., Chamerbo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.
 Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
 Nature 417 (6887), 459-463 (2002)
 22022145
 12024217
 2 (bases 1 to 10947)
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.B.A., Camarotte, G., Cannavan, F., Cardozo, J., Chamerbo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.
 Direct Submission
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
 FEATURES
 source
 Location/Qualifiers
 1. 10947
 /organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
 /mol_type="genomic DNA"
 /strain="ATCC 33913"
 /db_xref="ATCC:33913"
 /db_xref="taxon:190485"
 /note="pathovar: campestris"
 complement(85..501)
 /gene="XCC2657"
 complement(85..501)
 /gene="XCC2657"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 gene
 CDS
 complement(85..501)
 /gene="XCC2657"
 /note="synonym: XCC2661"
 complement(4847..5866)
 /gene="qxtB"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 gene
 CDS
 complement(4847..5866)
 /gene="qxtB"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

gene	<pre>/codon_start=1 /transl_table=11 /product="cytochrome D oxidase subunit B" /protein_id="AAM41933.1" /db_xref="GI:21113837" /translacion="MFVHLSVELMQHLSAAALFTVTVVVLVDGTLGVCMLEAANRS DADSHLWYSILTPVMDGNEWNLVGGSSILAMPEPSAYFLUNATYPLVVMALLALICR GVALLPYDQGLTDRRRVLVDVFWVASILATFCQGVVAGTYLAGVAHRAGYAGNGRE WLQFPFLFCGVALVGVLLGACGLWRSBQGLQQRHQAQRLVGLVIGALLTLGVWG LARUTPIYAGLQDAHSI.SVPIGTSVLGVMAMVLMRSHAFPLPLFAVLGVVATVATG VLAATYPMVLPPDITLQAASPPASQGMLLAFVLPCTLAYSYSYGFYVRGKVKPR GHAAE"</pre>
CDS	<pre>complement(5882. .7255) /gene="gxtA" /note="synonym: XCC2662" complement(5882. .7255) /gene="gxtA" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"</pre>
gene	<pre>/codon_start=1 /transl_table=11 /product="cytochrome D oxidase subunit A" /protein_id="AAM41934.1" /db_xref="GI:21113838" /translacion="MQASEHALLIVRGQATICTGMHIVLAAFTWGLANFLVLEALWL RTRRRVVDVNYVLMQSEHALLIVRGQATICTGMHIVLAAFTWGLANFLVLEALWL VLVAFLFLEAGFLGMLGKMKVGRHVHFATCMVALGSLVSAFWILSANSWLHTPAGF ARADGNFLPVDWLRILFNPSPPRMLHMSIALLATRAIIAAGVAHWLVRDTRNPGA RLMFSMGLWIAIVAPILQLLADGTLHGRLTLEHQPKLAIEGSNQRPAPGEGEPLFLF AIPMQQRNRHELIAPRLGSLYLRHNWTGTIAGLREFPRSDLPVPPIVFPFAPRMMV LGLLLIAGMLSLVLRAGRLYEARMQLQVYVVMFSGMLAMLCGMVWTEVGRQPTV YGLRLTAHSSTPLSLPMWASSVAILVLYPLVFLGLGLLRLVRRLPHARENGPSGG LADQTGN"</pre>
CDS	<pre>complement(7268. .7753) /gene="XCC2663" complement(7268. .7753) /gene="XCC2663" /note="putative; ORF located using Glimmer/Genemark"</pre>
gene	<pre>/codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAM41935.1" /db_xref="GI:21113839" /translacion="MAEPLDYDVKPPPIGFTAREBLDRLLTLHQRGILRLANDLGA QDKLAVLVGINKQPLGVNPSVILAMAVSRIPPSQFYKVFVACKGLADQIARSA QTAQGAEPGLRAYRLQLQDEAVNNAITPVLGDKLKAFAAGLDKQVDPKPI.SA9SKPSE S"</pre>
CDS	<pre>complement(7753. .10725) /gene="XCC2664" complement(7753. .10725) /gene="XCC2664" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"</pre>
gene	<pre>/codon_start=1 /transl_table=11 /product="formate dehydrogenase related protein" /protein_id="AAM41936.1" /db_xref="GI:21113840" /translacion="MQHSGSCTVYDPKTLHGQADIFLEIFLSGHGIDLSHCYHPAL</pre>

Alignment Scores:

Pred. No.:	2,63e-35	Length:	10947
Score:	749.00	Matches:	219
Percent Similarity:	43.02%	Conservative:	123
Best Local Similarity:	27.55%	Mismatches:	301
Query Match:	18.46%	Indels:	152
DB:	1	Gaps:	21

US-09-936-377-2 (1-758) x AE012378 (1-10947)

Qy	Qy	296	ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp	315
Db	Db	1521	---CCGGCGATCTGTCGATCGCGAGCCGGGTTTCG	1551
Qy	Qy	316	AspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAsnGlyLysArgTyr	335
Db	Db	1557	---CTGAAACTGCAGCAGCAGCCGTAC	1581
Qy	Qy	336	GluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuValHisLeu	355
Db	Db	1581	GACCTCAAGGGGGGTGACCGCATCGTGGGGCGATGGCAGCGCATTCGTTACAGTTC	1641

QY 356 AsnArgAsnAspTyrHisHisAspGluLeuAlaValGluAsnPhePheAsn 375
 Db 1641 GGCCATACCGGACACCGAATTCGAAGGCGCGAAGTCGCGACCGGTTCACC 1700
 QY 376 AsnGlnThrGlnAsnAlaGlgGluLeuArgHisGlnProIleGlyArgLeuTyrGly 395
 Db 1701 AAGCGCCCAAGAGGCGGTGTGAAGCTCTTACCTTCGCGCGCGCTGGCAGACT 1760
 QY 396 SerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
 Db 1761 GCCTTCGGCTGCAG-----GGCAGCGACACCCCTTCAGGCG 1799
 QY 416 Val----LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyVal 434
 Db 1800 GTGGCGAAGATCTTCGTGGCCCAAGACGATACCGTTCGTCGCGCTGTTCGCGCTG 1859
 QY 435 GluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLys 454
 Db 1860 GCCCGCAACAGCTGGGACCGCGTGACCGCGAAGTGGCGCGCGTGGACAAAGGTGAAG 1919
 QY 455 AlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeu 474
 Db 1920 TAGAC-----1925
 QY 475 ProAspLeuGlyAlaHisArgGln---ThrAlaArgSerPheAlaLeuSerGlyAsnTrp 493
 Db 1926 ACCGACATCGCGTGGACCGCGACTTCACCCCGACAGCTTCGCGCACGCGCGCTTC 1985
 QY 494 TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSer 513
 Db 1986 CGCTTCACGAGCAGTGGCGCTCACCAGCCACCTCGACCATCGCGACGCGCGCGGCC 2045
 QY 514 ThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsn 533
 Db 2046 GAGCAAGAACTGTTTGGCAACGCGCGCACATCGCCACCTCGACCATCGCGCGCGCGCAC 2105
 QY 534 LysHisLeuAsnLysGluArgSerAspAsnIleGluLeuAlaLeuGlyTyrGluGlyAsp 553
 Db 2106 GCCAATCTCAAGACGAAGAGCCCAACGAGCGCGAGCTGGTGTGAATTTCCAGAACGAC 2165
 QY 554 ArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr----- 571
 Db 2166 TGGGTGGATGCAAGATCGCGCGTACTACACCGCTACACGAGCTTCGCTATATCGTC 2225
 QY 572 -----AlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSer 588
 Db 2226 GATACCGCGCGCGAGTGTTCACAGAA-----GAAGACACGACG 2264
 QY 589 GluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyValGluGlyGlu 608
 Db 2265 TTCCTGCGGATCGCGCAGTGGACGCGCAGCGCCATCTCCACGCGTTTGAAGCGAA 2324
 QY 609 IleTyrPhe-----LysProThrProArgTyrArgIleGlyValSerGlyAsp 624
 Db 2325 GCCACCTTCACCTCGGCAACACGACCGCGCGATGGGACCTTCGCGCTTCGCGGAC 2384
 QY 625 TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly 644
 Db 2385 ACGGTGCGCGCGCTGAG-----2405
 QY 645 AsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeu 664
 Db 2406 -----GATGGCGCACTTCGCGCGCATCGCACCGCGCGCGGTG 2444
 QY 665 GlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArg 684
 Db 2445 GCGCGGGAATCGCTGGAC---GCCGATGATCGCGCGCGTTCGCGCGCGCATCGCG 2501
 QY 685 ValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeu 704
 Db 2502 ACCATGAAGCAGGACAAAGTGGCGGTGAAGCAACCGCTACCGACGCGTACACCATGTG 2561
 QY 705 AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsnTrpAsnTyrValLys 724

Db 2562 GACGCGCATCTGGCTTACCACATGATCGCGCGGCCAACGCGTGGGAAGTGTCTCTCGAT 2621
 QY 725 AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro 744
 Db 2622 GGCACACACCTTCCACCAACAGGATGCGCGCGTGCACACCTCTCTCTCAAGGACGAGTG 2681
 QY 745 GlnMet---GlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2682 ATGTGCGCGCGCGCAACGCTCTCTGCGCGTGGCGCTGTCTTC 2726
 RESULT 18
 LOCUS AE011924
 DEFINITION Xanthomonas axonopodis pv. citri str. 306, linear BCT 29-MAY-2002
 the complete genome.
 ACCESSION AE011924 AE008923
 VERSION AE011924.1 GI:21109120
 KEYWORDS
 SOURCE
 ORGANISM
 Xanthomonas axonopodis pv. citri str. 306
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 Xanthomonadaceae; Xanthomonas.
 REFERENCE
 1 (bases 1 to 11947)
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,
 Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
 Ferreira, R.C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Greggio, C.C., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
 Madeira, A.M.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.P.,
 Taktai, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
 Kitajima, J.P.
 Comparison of the genomes of two Xanthomonas pathogens with
 differing host specificities
 Nature 417 (6887), 459-463 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 22022145
 12024217
 REFERENCE
 2 (bases 1 to 11947)
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,
 Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
 Ferreira, R.C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
 Madeira, A.M.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.P.,
 Taktai, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
 Kitajima, J.P.
 Direct Submission
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
 Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
 Brazil
 JOURNAL
 FEATURES
 source
 Location/Qualifiers
 1. 11947
 /organism="Xanthomonas axonopodis pv. citri str. 306"
 /mol_type="genomic DNA"
 /strain="306"
 /db_xref="taxon:190486"
 /note="pathovar: citri"

Query Match:	17.80%	Indels:	149
DB:	1	Gaps:	21
US-09-936-377-2 (1-758) x AB011924 (1-11947)			
Qy	11	LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAla-	24
Db	1333	CTGGCCCTGGCAGCCTGCTACCCCTGCTCCCTGGCCATGGCGGAGGACGCTCCGCTCGCT	1392
Qy	25	-----HisGluThrGluGlnSerValGly 32	
Db	1393	GCCGATGGGCAGACCCCGCCAGTTTCGACACACGCCAGGAGGAGCCCTCCAGTGGG	1452
Qy	33	-----LeuGluThrValThrValValGlyLysSerArgProArgAlaThr 47	
Db	1453	CGCCATATCAAGGATCTGGACAAAGTG--GTGGTCACCGCAGCCCGTTCGCGGATGCG	1509
Qy	48	SerGlyLeuLeuHisThrSerThrAlaSerAspLysIleLeuSerGlyAspThrLeuArg 67	
Db	1510	GCCGTGAGTTGAGCGCCCGCGTG-----GAAGTGTGGCGCGGAGCGTCTGGAC	1560
Qy	68	Gln---LysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86	
Db	1561	GAGTGGCAGTTCAGCTGGCGGAAACCGTGGCCAGCTGCGCGGTACAGAGTTC	1620
Qy	87	GlnThrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLys 106	
Db	1621	AACCTTCGGCGCGCGGTGGCGCGCGGATCATCCCGCGCTGGAGCGCGCGGTGGCA	1680
Qy	107	ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126	
Db	1681	GTGCTGGCGCGCGCTGTGACCCAGACGCTGTCACCGTCAGCAGGACCATCCCG	1740
Qy	127	MetValAspThrAlaLeuSerGlnValGluIleLeuArgGlyProValThrLeuLeu 146	
Db	1741	GCAATCGAGCGCTTCTCGCCAAACAGATCGAAGTGTCTGAAGGTCCATCGACGCTGCT	1800
Qy	147	TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166	
Db	1801	TACGCTTCGGCGCATCGCGCGGTGTCACTGCTGACGCGCGCATCGCGAG--	1857
Qy	167	MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186	
Db	1858	ACCCCGGTGGAGCGGTTCAGCGCGCGCGCGGAGTACGCTTCGACGCGCGGACAAAGGAC	1917
Qy	187	LysLeuThrSer---GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThr 205	
Db	1918	GGCAACACCGACATGTTCCGCTGATCGCGCAATGGCAGCGCGCTGTCGATCCAGCA	1977
Qy	206	GluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225	
Db	1978	GATGGCGTGTATCGCAACAGAACGACTACGACACCGCGGAGGGCGCGAGCTCAACAGC	2037
Qy	226	LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArg 245	
Db	2038	TGGGTTCGATTC-----AAGTGGCTCGATCGGTCGCTGCTCCGCGCAT	2085
Qy	246	LysArgPheTyrArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHis 265	
Db	2086	TGGGTTCGTTGTTGCGCATCGCTTTCGCGATTAACCTACGCGCAACCGCGCGAG	2145
Qy	266	SerHisGluTyrAspCysHisAlaAlaIleIleThrGlnLysSerLeuIleAsnLys 285	
Db	2145	-----	2145
Qy	286	ArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnPro 305	
Db	2146	-----CCGGCGGACCCCTCCATCGCGGAGCGC	2172
Qy	306	GlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGlyLys 325	
Db	2173	GGGGTGTG-----	2181
Qy	326	ProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPhe 345	
Db	2182	-----CTGAAGTTGCAGCAGGACCGCTACGACCTCAAGGCGCGGTGACCGATCCGTGG	2235
Qy	346	ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLys 365	
Db	2236	GGGGAAGGCGAGCGCGTGTACAGCTTCGCGCATACCGATTACCGCATCCGAATTC	2295
Qy	366	AlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeu 385	
Db	2296	GAGAGCGCCGAAGTCCGACCGCTGTTCACCAAGCGGCGCAACGAAGCGCGTGGAGGCC	2355
Qy	386	ArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLys 405	
Db	2356	TCGTTTACGTTTGGTGGCGCTGCGAGATGCAATCGCTCGCTGCG--	2400
Qy	406	SerSerAlaLeuSerAlaThrSerGluAlaVal---LysGlnProMetLeuLeuAspAsn 424	
Db	2401	-----GGCAGCGACAGCACCTTTCAGCGCATCGCGCAAGAAATCTTCGTGCGCAAGACC	2454
Qy	425	LysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyrAspAsnPheThrLeu 444	
Db	2455	GATACAGCTCGCTCGCGCATTCGCGTTCGACGCAATAGTTGGAGCGCGTCACCGCC	2514
Qy	445	GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLe 464	
Db	2515	GAAGTCGGCGCGCGGTGGACAAG-----GTCAAGTACCAG--	2550
Qy	465	AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---Thr 483	
Db	2551	-----ACGATATCGCGCTGCGACCGCTGATTTCATCT	2580
Qy	484	AlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeu 503	
Db	2581	CCGACAGCTCTCGCGCAGCGCGTTCGCTTCAACGAAATGGCGTCTGACCGCC	2640
Qy	504	ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523	
Db	2641	AACCTGACACCGCGCGCGTGGCGCGCGAAGAAAGAACTGTTCGCAACCGCGCGCAT	2700
Qy	524	ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543	
Db	2701	ATCGCACCTCGCTTCGAAATCGCGATGCAATCTGAAGACCGGAGAGGCCAACCCAG	2760
Qy	544	IleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg 563	
Db	2761	GCGAGCTGGCTTGAATTTCCAGAACGATGGGTGGATGCGCAAGTGGCGCGGTACTAC	2820
Qy	564	AsnArgPheGlyAsnTyrIleTyr-----AlaGlnThrLeuAsnAspGly 578	
Db	2821	AACCGCTACCAACGACTTCGTCTACATCGTCGACACCGCGCGCGCTGTTCAACGAA--	2877
Qy	579	ArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSer 598	
Db	2878	-----GAAGACAACGACTTCCTCGCATCCGCGAGTGGAGCGAGGCC	2919
Qy	599	GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe-----LysProThr 614	
Db	2920	GATCGCATCTTCACCGTTTCGAAGCGAGCGCCACCTTCACCTGCGCAACAAAGATAC	2979
Qy	615	ProArgTyrIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuPro 634	
Db	2980	GGTGGCTGGACCTCGCGGTTCGCGCATACGTTGCGTGGCGGCTCTG-----	3027
Qy	635	SerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGln 654	
Db	3028	-----AACGATGCGCGC	3039
Qy	655	AsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp 674	
Db	3040	AACCTCGCGGTATCCACCGCGGACGCTTCGTTGCGCGCATGCGCTGGGAT---GCCGAC	3096
Qy	675	ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyr 694	

```

Db      3097  CAGTGGCGCATCGCTTGGTCCACCGCACCCTGAGCAGGCAAGGTGGCAGTGAAC 3156
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      695  GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThr 714
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      3157  GAAACCCCGCAGCGATACACCTTCCTGATCGATCGCATCTGGCGTATCATCGTGGACAC 3216
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      715  ArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyr 734
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      3217  GCGACCGCGCGTGGGAAGTGTCTCTGCGACGCAATAACCTGACCAATCAGGACGCGC 3276
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      735  AlaHisSerSerPheLeuSerAspThrProGlnMet---GlyArgSerPheThrGlyGly 753
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      3277  GTGCACACCTCTCTCCTCAAGGACGATGATGATCTCTGCGCGCAGCGCTTCGTTCGGC 3336
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      754  ValAsnValLysPhe 758
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      3337  GTGCGCTGTGTTCTTC 3351
               |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 19
AE003890
LOCUS
DEFINITION
Xylella fastidiosa 9a5c, section 36 of 229 of the complete genome.
ACCESSION
AE003890
VERSION
AE003890.1 GI:9105215
KEYWORDS
Xylella fastidiosa 9a5c
SOURCE
Xylella fastidiosa 9a5c
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xylella.
1 (bases 1 to 10158)
Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,
Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furian,L.R.,
Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
and Marino,C.L.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
Nature 406 (6792), 151-157 (2000)
20365717
10910347
2 (bases 1 to 10158)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furian,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuranae,E.E., Laigret,P.,
Lambais,M.R., Leite,L.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani,Jr., A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,A.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,

```

```

TITLE
Xylella fastidiosa 9a5c, section 36 of 229 of the complete genome.
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furian,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuranae,E.E., Laigret,P.,
Lambais,M.R., Leite,L.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani,Jr., A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,A.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,

```

```

Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil
FEATURES
source
1. .10158
/organism="Xylella fastidiosa 9a5c"
/mol_type="genomic DNA"
/db_xref="taxon:160492"
/clone="9a5c"
/complement(302. .484)
/gene="XF0382"
/complement(302. .484)
/gene="XF0382"
/notes="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83192.1"
/db_xref="GI:9105216"
/translations="MGKGRKTTKGRKYNASYNARPHSISKAPVDAAPLNNRRAPK
AQSPRTVVKTVAKAI"
/complement(609. .1088)
/gene="XF0383"
/complement(609. .1088)
/gene="XF0383"
/notes="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83193.1"
/db_xref="GI:9105217"
/translations="MPSNSKVLHYSKLLMLFKNIKAMLFTRHRLDHPGATGSLIC
ALHCAPLPLLVIAAPAGPFLWFGDVGQVIVSVTLALFSLISYFKHRTWRAISI
LLGLTLLMLGLLYDPLHAPTHLAVTMAIGTTLGIAHLVNLHSHIDNTSSKH"
1175. .3220
/gene="XF0384"
1175. .3220
/gene="XF0384"
/notes="similar to GI|3044098 (percent identity: 22 %/query
alignment coverage: 88.3 %/subject alignment coverage:
78.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="outer membrane hemin receptor"
/protein_id="AAF83194.1"
/db_xref="GI:9105218"
/translations="MLVTHISILPRTVFLVAVSSLLTTPALAMADVCPPTDSRHFKT
TEVTRHUKLDADVVTALPDRSDYSLSRVPVALLAGERLDEVRASSIGETIVAVLPGVQ
SSNPGVGPIIRGLDGPRIAVLNGLSQDVSQDSHPAVEPFLANQIEVLKGVQ
STLLYGSAIGYGVNIVDGRIAEAPVGVGGRAMRLDGGDKGNTNMFRIADAGSA
LHSHADGVYNEKDYDTPKQVNVNSFTDKSGSMGASFGDFGFLSVARFHSYGN
PGEQDPAVAGRSWLRHODRYDLKAGLTPWGEASALRFSLHTCYDHIPEFGEV
TTPGKASGRVSEAFPGGWRTPAGVGGSTFQALGESEFPKYNKNSIGVGL
AHNTFGPQAFPGARQDVQITDINGVTRNHFSGLSAFSGDLALSQKWRLLTNVDHAE
RNPVEELFGPHATLAYEVRQDLKKEKQAEGLVFRNDWSDAKYSTYSRYG
NFYLVDTGTFWDDQRLPVQMSQANAFHGEATFHLAKNTSGSWDLRVFG
SGYGRKNGNLPRIVPARYGAELWEDAGWRTSLSAKYEKQNVAVETPTAGYT
MVDAHLAYHIDVDSIAWEVFDGNLTNRDVRHTSLKVDVMLAGRNYTAGLRMPF"
complement(3759. .4127)
/gene="XF0385"
complement(3759. .4127)
/gene="XF0385"
/notes="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```



```

/transl_table=11
/product="DUF190"
/protein_id="CAD85614.1"
/db_xref="GI:30180702"
/translation="MNGYQITFTTQDRHHAGKPLADWLMLHAAEMGLGATLIPGSE
GMGDKRPHSVHFELSDQLLEVVMVVEEADQLFARLEAGVRLFYVYKVAEFGIM
D"
gene
complement(454..837)
/locus_tag="NEI1704"
CDS
complement(454..837)
/locus_tag="NEI1704"
/function="InterPro IPR003691 COGs COG0239"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CRCB protein"
/protein_id="CAD85615.1"
/db_xref="GI:30180703"
/translation="MKKPIALALGSTLGGLLRWGLGLKLNLPDVPPTGLVANLIA
GYVGVAFAPFAHNPENSPWRLLVITFGCGGSTFTFSARIIVSLLRGLYAWASA
IAVHVAGSLIMTLAGIATVWFKSS"
misc_feature
order(complement(769..825), complement(664..729),
complement(568..633), complement(466..531))
/locus_tag="NEI1704"
/translation="InterPro IPR00432:IPR001687:IPR002863 COGs
COG0249"
/evidence=not experimental
/transl_table=11
/product="mutS: DNA mismatch repair protein"
/protein_id="CAD85616.1"
/db_xref="GI:30180704"
/translation="MNKAQSSHTPMQOYLRIKAQHTDKLFLVRMGDFVELFYEDAE
KAAKLDTLQRGSGAGEPIKMAQVPPHAPPAADQYLARVLGSGIAICEQGDPAISK
GPVERQVIRLTPTGLTAGLLEERSNIVLALALHRSIGLAWLNLAAGDMRWLETS
SDNLTSELERLHPALLESIDLPATLNNFAGPKRLPDWOFDVEHAMQOLTROFGTR
DLNARCCDLHRAIMAGALREYVRLTQCTATDGSSQGLPGHLHTLOVERQDAVLMD
AATRNLEITLIRGEDAPTLSLLDCTGSGKRLHRLHRLNRIILQQLRDYV
SDLIQAQETTYAGIRQOKPHADIERITSRILAKPARDLSGLRSLMKPLGILSL
IATSAARAVHRIPIPMQPDPLLTQLLVRALQVPVAGVIREGGVIADGFDALDELRLG
QNCDEPTLQLEAREERTGIPNLKVEYNRVHGFYIEVTRAQGEKIIPDYRRRLTKN
AERYIIPELQAFEHKTLASAREALARKMLYERLIEQLADFIIPLOIARSVAELDVL
CAFARAALSGYTKPVTDDPVLITTEAGRHYPVENQVEHYIANDVOLCAITRENQOML
VITGPMGKSTYMRQTMVLVLAHCGSFVPAQIARIGPIDQIFRIGAADLAGGRS
TWVMTERRAGTIRNATAGSLVLDVEIGRGTSTFDGLALAPAIARHLLTONQSYLFA
THYFELTRIAEFPQAVNIHVTAVEHKRIIVFLHRIEGSPASRSYGLHVAALAGVPR
VIRNAAKILARDEQTLRSPOQTLTFETVEENAKAVPASVHPVLDYLERIHPDELTPR
GALEQLYLKSMNLQTD"
gene
complement(3899..4378)
/locus_tag="NEI1706"
/translation="InterPro IPR001179 COGs COG1047"
/evidence=not experimental
/transl_table=11
/product="FKBP-type peptidyl-prolyl cis-trans isomerase
(pPIase)"
/protein_id="CAD85617.1"
/db_xref="GI:30180705"
/translation="WRIVKNTVVVSLHYKIMDOEGNMWETPEFVSYLHGGYDGIFFPAV
EELALHEKETGAALSQMEPDAFGEYDELMVEVPNAPPEVA/GMQFGEDESDSD

```

```

gene
complement(4393..5046)
/locus_tag="NEI1707"
CDS
complement(4393..5046)
/locus_tag="NEI1707"
/EC_number="3.1.26.4"
/function="InterPro IPR001352 COGs COG0164"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Ribonuclease HII and HIII"
/protein_id="CAD85618.1"
/db_xref="GI:30180706"
/translation="MAERRIPLKHEYAODGKVIYGVDEAGRGLAGPVYAAACVLDPA
DVIEGLADSKOLSEKKRISLADQIKORARAWAIASVVEIDRLNLIQASLLAMQRAV
VSLRPSNALVLVDGNHAPRLDCEVQTVIRGDSLVAEISAASIIAKTARDIEMLRLEH
AYPVGPRHKGYPYKAHLEAIRLGIITDIHRRSFAPCVGQSVSGARTTTSFNOKEA"
complement(5068..5529)
/locus_tag="NEI1708"
/EC_number="4.2.1.1"
/function="InterPro IPR001143 COGs COG0764"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Bacterial thioester dehydrase"
/protein_id="CAD85619.1"
/db_xref="GI:30180707"
/translation="MKQGRQTVMDIHEILKYLPHRYPLILVDRVSVLSGKRIHAYK
NWSINEPYFSGHPHPHVPMPGLIIEALAAALITIRSENMEKONGQVYVFGIDAV
RPKPVIAGDQVLKVEITRQLGKWKYACAAEVDGVVTEAQLMCTARAI"
gene
complement(5546..6109)
/locus_tag="NEI1709"
CDS
complement(5546..6109)
/locus_tag="NEI1709"
/function="COGs COG2825"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative transmembrane protein"
/protein_id="CAD85620.1"
/db_xref="GI:30180708"
/translation="MRIPDQYQWQGFRRFVAFVVTMMFVLVPHSSAGEIKIGVNT
EKLVSRSQYQRAQQMRDLSLRQNEEYGLIERINQVIRELAEKQSYDLILQLODS
VYRSARIDITDQVTKVNARESAARKP"
misc_feature
complement(6005..6070)
/locus_tag="NEI1709"
/translation="InterPro IPR000184 COGs COG0729"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Bacterial surface antigen (D15)"
/protein_id="CAD85621.1"
/db_xref="GI:30180709"
/translation="MKLRLLIFFLSLYSLGCMANDSLVVRDIVERGIQTEAGTVESY
LPVKGVLDLSKASAAIKALYATFFSDVKLSEGLLIIVQWERPALAQISNGAK
EFPDKLKEGLKQAGLSERYSFISRLLEKAEDELKQYISRGKVAVKITTTTFLERN
RIGINFDIKEGKTARIKQINVGNIWFFPEDDLVDLFSLKTPGFWFTTKDDQYSKQKL
SADLETLSRVDLYLENIESTQVITPMKDIYITVNTVEGTPQYVSDIKLAGEL
LVPEEELRKLILEPGGIFVREKLITSEIKLISDLRQNDGVAFANVNASPELDEKTKT
AFTFFIDPGRVYVRINISGNERTDEIVRFRQMEGWHSTEQINRSRQRVDLELD

```



```

Db      18006  CTGCCAGATCGGTCATATGAGGACGG-----|||||
Qy      590  MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu--- 608
Db      17961  TATTTGTTATTGAAGTACCGTCMAAATAATGCAAAATTTCTCGGCGATGATTTGAGACG 17902
Qy      609  -----IleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyr 625
Db      17901  ATTGTCATCTGCTCGAGTGATCGTCGGGTAAATTTGGATCTCGCTTGTGGACTGATAT 17842
Qy      626  ValArgGlyArgLeuLysAsnLeuProSerLeuPro-----Gly 638
Db      17841  GTCCGCGGGCGCTGTCATCAGGAGGATATTACCGCGCATGACACCATTTTCGTTTGG 17782
Qy      639  ArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArg 658
Db      17781  GGGGCACTTGATATTATGCTACCGTCCATGG----- 17752
Qy      659  ValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAla 678
Db      17751  -----CGAGGT 17746
Qy      679  AsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThr 698
Db      17745  CGGATTCAGCTAATGCGGGTGCATTAACAGCGCGATATCGCTCTGCTCGAAACGACACC 17686
Qy      699  ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu 718
Db      17685  GCGGTTACACTACTGCTGAATGTTCACTGGATACCGTTGACCTGGGGAAGATGAT 17626
Qy      719  TrpAsnTrpTyrValLysAlaAspAsnLeuLysGlnSerValTyrAlaHisSerSer 738
Db      17625  TACAACCTGTCATACGAGTGCATATCTGCTGAATGAGAGCGGAGACGCGCACACTTCT 17566
Qy      739  PheLeuSerAsp---ThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLys 757
Db      17565  TTTTGAAGATCGGCGCACCGTTCCTCGGCTCGGCTGGTGGAGTGGCGGTGAT 17506
Qy      758  Phe 758
Db      17505  TTT 17503

RESULT 21
AE012559/c
LOCUS   AE012559.1 300885 bp DNA linear BCT 29-JAN-2003
DEFINITION Xylella fastidiosa Temeculal, section 7 of 9 of the complete genome.
ACCESSION AE012559.1 GI:28057550
VERSION   AE012559.1
KEYWORDS  Xylella fastidiosa Temeculal
SOURCE    Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xylella.
ORGANISM  Xylella fastidiosa Temeculal
REFERENCE 1 (bases 1 to 300885)
AUTHORS  Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorri, H., Tsai, S.M., Carrier, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Blanco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Penille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., de Oliveira, A.R., Rosa, Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P.
TITLE     Comparative Analyses of the Complete Genome Sequences of Pierce's Disease and Citrus Variegated Chlorosis Strains of Xylella fastidiosa

```

```

J. Bacteriol. 185 (3), 1018-1026 (2003)
12533478
2 (bases 1 to 300885)
Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorri, H., Tsai, S.M., Carrier, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Blanco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Penille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., de Oliveira, A.R., Rosa, Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P.
Direct Submission
Submitted (17-DEC-2001) Instituto de Biociencias, Universidade de Sao Paulo, Rua do Matao, 277, Sao Paulo, SP 05508-900, Brazil
Location/Qualifiers
1. 300885
/organism="Xylella fastidiosa Temeculal"
/mol_type="genomic DNA"
/strain="Temeculal"
/db_xref="taxon:183190"
/notes="Pierce's disease strain"
87. .1022
/gene="apbE"
/locus_tag="PD1557"
87. .1022
/gene="apbE"
/locus_tag="PD1557"
/locus_tag="PD1557"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="thiamine biosynthesis lipoprotein ApbE precursor"
/protein_id="AAO29399.1"
/db_xref="GI:28057551"
/transl_table=11
/translation="MGTTWSVKLAAPRRDLPLHNAIQARLDQIVAEKSTWSPNSHI SRNPAPVRSWHPLEDFDIATLDIANTSGADPTVGLVRLWGFAGVSGNHKO IPTPEATLATRSGQHLETTDGHWLOPGTDELDSATAGKGYSDAAVATHGKQ EHALLDGCGLYGHGKPDGTWNSVLVEMDLTHQYGNPLPFCILQDLGLAVATSGDRW HPEHEGQRVTHTDHPQGTPIIPALPVTMIASSAHADAWATALTILGREAGLALA NTGLAVRYLEHNDAFTAYYSPAFTELLHKPATA"
/gene="comE"
/locus_tag="PD1558"
/complement(1752..2081)
/gene="comE"
/locus_tag="PD1558"
/locus_tag="PD1558"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="DNA transport competence protein"
/protein_id="AAO29400.1"
/db_xref="GI:28057552"
/transl_table=11
/translation="MKSFAMLKSLLLFLVLAFTVAANKVDINTASAEEMDKVLVNI GPSKASAIKVRRENGSPFKSVEELALVKGICMKTVERNRDLLETGTRMAPAKHAKGTE LKSVGRK"
2603. .3625
/locus_tag="PD1559"
2603. .3625
/locus_tag="PD1559"
/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO29401.1"

```


QY	156	ValaspValalaaspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGlu	175	Db	148013	TTGAACTGTCATGCGGAGCGTGTCCGGTGGAGGAGAACTGTTTGCACAGGCGCG	147954
Db	148823	GTCATATCTGTATGAGCGATATCTGTA---GGCCCGTAGGTGGTTTCAACGGAGT	148767	QY	523	HisValalaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAspSerAsn	542
QY	176	LeuGlyLeuLeuSerSerGly-----AsnLeuGluLysLeuThr	189	Db	147953	CACATCCGACTCTTGTATGAAGTTGGTCGCCGACCTGAATAGGAGAGGCTAAC	147894
Db	148766	GCTGAATAGCTTTAGACGGTGGTGATATAACATGTAATAACAACATGTCGCAATTGAT	148707	QY	543	AsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrPheGlnTyrAsnLeuAlaLeuTyr	562
QY	190	SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyr	209	Db	147893	CAAGCTGAGTTGGGTTGGTATTCCGTAACGATTGGAGTGCATCCAAAGTTTCGACCTAT	147834
Db	148706	CGTGGTAAC-----GGAGTGTCTATTCGTCATCGCAGATGCCGTATAC	148662	QY	563	ArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyAlaGlyProLys	582
QY	210	ArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSer	229	Db	147833	TACAGTCGCTATGCAATTTTCATTAC-----TTAGTCGAT---ACAGGAGTACT	147786
Db	148661	CGTAACGAAAGGACTATCACACCCGAAAGGCGCGCAGGTAACTCGCTTCATCAGACACC	148602	QY	583	SerIleGluAspSerGluMetLysLeuVal-----ArgTyrAsnGlnSerGlyAla	600
QY	230	ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyr	249	Db	147785	TGGTTCGAGTATGACAGCGGTGATTTCGCGTCCGTCAGTGGTCGCAAGCTAATGCA	147726
Db	148601	-----AAATCAGCTCGATGGGTGCTTCGTCAGCGGTGATTTCGTTCTTT	148554	QY	601	AspPheTyrGlyAlaGluGlyGluIleTyrPhe-----LysProThrProArg	616
QY	250	ArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyr	269	Db	147725	ATTTTTCACGGAATTGAAGGTGAGCCACCTTTTCACTTGGCTAAAAAATACGAGTGGTCT	147666
Db	148553	GGCTGTGCGTCGCGGCTTTTCATGACAGCTATGGC-----	148518	QY	617	TyrArgIleGlyValSerGlyAspTyrValArgGlyValGlyLeuLysAsnLeuProSerLeu	636
QY	270	AspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGln	289	Db	147665	TGGGACCTACGTGTGTGGTGACGGGTTCGTGGTCTTTGAAA-----	147621
Db	148518	-----	148518	QY	637	ProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAla	656
QY	290	LeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys	309	Db	147620	-----GATGGAGGCAATTG	147606
Db	148517	-----AATCGGGT-----	148509	QY	657	ProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIle	676
QY	310	GlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAsp	329	Db	147605	CCAGCATCGTTCAGCACGCTTACGGGCGGAGGTACGTTGGGAG-----	147561
Db	148508	-----GAGCTGTGTATCTATCCGCGGTGACCGCGGTAGTGGTGTCTA	148464	QY	677	Asp-----AlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla	692
QY	330	LeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheGlyPheGlu	349	Db	147560	GACGTTGGTTCGGCTACATCGTTGGTGTCTAAGCGCTACGAGAGACAGACAGAGTAGCG	147501
Db	148463	TTGCATCAGGATCGTTATGATTGAAGCTGGTTAACTGATCCATGGGTGAGCGGAGT	148404	QY	693	ArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArg	712
QY	350	AlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAla	369	Db	147500	GTCACGAGAGATCCCAACACAGGATATACCATGTAGATGCGCATTTGGCTACCATATT	147441
Db	148403	CCCTCGCTTCAGCTAGCTACACTCATATATGATGATCATATTGATTGAGGGAACGAG	148344	QY	713	AsnThrArgTyrGlyGluTrpAsnTyrValLysAlaAspAsnLeuLeuAsnGlnSer	732
QY	370	ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro	389	Db	147440	GATGGGATAACATGCTGGAGGTCTTTTCATGGTAACATCTCCACCACTGCGAT	147381
Db	148343	GTCGGCACTACTTTCGTAAGCGTCTTCAGAGGGCGAGTCCAGGCTCTTTCGATT	148284	QY	733	ValTyrAlaHisSerSerPheLeuSerAspThrProGlnMet---GlyArgSerPheThr	751
QY	390	IleGlyArgLeuLysGlySerTrpGlyVal-----GlnTyrLeu	402	Db	147380	GCTCGTGTTCATCTCTCTTAAAGACGATGTCATGTGTGGTGGTTCGTAATACTACT	147321
Db	148283	GGTGGTGGCTGGCGGACTGGTGGTGTGTCAGGGGAGGACTCTACTTTCAGGCGCTT	148224	QY	752	GlyGlyValAsnValLysPhe	758
QY	403	GlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeu	422	Db	147320	GTTGGTTTCGATGTTCTTT	147300
Db	148223	GGTGAGGAGTCATTTGTACCAAGACGAATAACAATCAATT-----	148182	RESULT 22			
QY	423	AspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspPhe	442	AE004513/2			
Db	148181	-----GGTGTCTTCGATTTGSCATACATCACTTTTGGGTTGTTT	148143	LOCUS			
QY	443	ThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAla	462	DEFINITION			
Db	148142	CAGTAGAATTTGGTGGCGGTGGTATCA-----GTTAAATACGATACGGAT	148095	ACCESSION			
QY	463	LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln	482	VERSION			
Db	148094	AATGGCGTTACTCGGAATTATCAC-----	148071	KEYWORDS			
QY	483	ThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer	502	SOURCE			
Db	148070	---CCAGGAAGTCTTTTCAATTAGTGGGTATTTAGCACTCTCTAAGCAATGCGCGTACT	148014	ORGANISM			
QY	503	LeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys	522	REFERENCE			
				AUTHORS			

AE004513 15470 bp DNA linear BCT 19-FEB-2003
Pseudomonas aeruginosa PA01, section 74 of 529 of the complete genome.

AE004513 AE004091

AE004513.1 GI:9946669

Pseudomonas aeruginosa PA01

Pseudomonas aeruginosa PA01

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 15470)

Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warriner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,

Wong, G.K.-S., Wu, Z. and Paulsen, I.T.
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
10984043
2 (bases 1 to 15470)
Stover, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.D., Coulter, S.N., Folger, K.R., Kas, A., Iarbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 15470)
Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
Direct Submission
Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada

This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from PseudoCAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.
Class 1: Function experimentally demonstrated in *P. aeruginosa*.
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

Location/Qualifiers
1. .15470
/organism="Pseudomonas aeruginosa PA01"
/mol_type="genomic DNA"
/strain="PA01"
/db_xref="taxon:208964"
126. .878
/gene="pruR"
/locus_tag="PA0780"
126. .878
/gene="pruR"
/locus_tag="PA0780"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in *P. aeruginosa*); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 2; Alt protein name: probable transcriptional regulator"
/codon_start=1
/transl_table=11
/product="proline utilization regulator"
/protein_id="AAG04169.1"
/db_xref="GI:9946670"
MRACLVPGEAGHVFAGVDNRMILIDGENDPTDPDLARLFEAPRYPTLDAOFN
LLSVAGELARVSDLLTRALGGVLRALRLFGFQPLASGPDLLERLDAHVR
NIARRITVAELAAQACUSPHFQAQKDSVGLTPTHTYLLKVRLLDCAARLLRESRLPLV
RIARECGSSSALTTAWRRYLGITPKSLRG"

gene
CDS

complement (999. .3062)
/locus_tag="PA0781"
complement (999. .3062)
/locus_tag="PA0781"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences); Subcellular localization: Outer membrane protein; Subcellular localization confidence: Class 2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04170.1"
/db_xref="GI:9946671"
/translation="MSSSLPSPRLPMLPTPLALACLIVSGETLGGADGRSELPSQVITANPLGNESPATPSSVLEGGDELTLROKSLGETLNGLPVGSSTVFGPGASRPVIRGMDGDRFLNGVGDALDASSLSYDHAHPEDPNSVERLEVGRPAALLYGVNAIGGVNSFDNRIPSEFVDGIHSGELRYGGADUTRSRSGALEAGDGNFALHYDAASRENDVRIFGVAHSRQRQIDGDTGKHVRQNSDQGGAVGSGSYHHEHAGLSISYGTDSNYGSPAE DDVRLKMQDDRYAFASERIDLEGPTSLKLDAAVTKYHEKEIDEGTSTFKNGYEGRIEARHRLPLGVLNGVGAQFANRFSALGEEAFVPHETDSAAALFALEWKLSDRLDL SFGARLEHTRVDPDAKGNRFAENDQSFTTSLSTGAVYKLTPIWLSAATLSYTERAPTVELYANGPHATGTYEVDADADKEKAVSTDALRFDNGVHKHSGVGVFSFNS YIGLLASGRHNEGEVVAAGDDELPEYLYSGVRADFYGEARHILHLSFSPFNFD LELSGDYTRAKNKOTGEPLPRIAPRLNTALILWELQQAQVDDVEHAASQHRVPEBEL STDGYTTIGASLIGNFDIGESRWLAFVKGTLNQTIVRYASSILKDRVPAAGRGIEAG VKVAF"
3526. .6708
/gene="putA"
/locus_tag="PA0782"
/note="synonyms: prUA, prUB"
3526. .6708
/gene="putA"
/locus_tag="PA0782"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in *P. aeruginosa*)"
/codon_start=1
/transl_table=11
/product="proline dehydrogenase PutA"
/protein_id="AAG04171.1"
/db_xref="GI:9946672"
/translation="MFKASHVLQGDQTRSMAEFPFVISAANYVDDEAVLGLLQALD PGDGIERARNARALIFTRSRDNADVTLDALRQYSLDTQEGMLMCLAEALLRVP DAATDALIRDKLSAADQWRLHQCSDNVLVPAWGLVMTGKVNIDERTGGRASVIV SLEVRSGEPYIRAAWQAKMLNGQFVIGRTISLAKNGRPCRQGTYYISFDMLGEEA ALTAADAKIMADYIQADITVGAEPFGVGPSPSISIKLSALHPRIEVARQRVUSGEL FANVLELAVRAKLVNGITIDAEADRLLESLLEYKLMRDPAIAGWGEFGGLVIQAYS KRCPLVYMLTLGLRELGARIPRLVKGAYWDSIKQVQGDADGYPIYTRKEGTDTS YLACARYLISEHTRGVIVYQFASHNAHTVTAIALADARAAGSSEFVHKLVDPVPVETL DALYDTITEKYRNRVRIYAPGAHKDLPYLVRLLENGANSFVHKLVDPVPVETL TOHPVTQROFCTLANDRIPLPOALFGAGKNSQGINNIONWNELELAYKPHLTRQ WQARIVNGEKLSPAHVRCPYDRLERLVGSVQYASAEQAKALDVLQAWPRNATPQ VDERAGILERLADLLEAGQRMALCTLEAGSLQGDIDDEIEADFCRYAQQAQRQ LVREELRGPTGERNELFYEGRGIFACVSPWNPFLAI FLGIQSAALVAGNTVLAKPAEQ TSLIAARTVELMFEAGLPKDVIALPGDGTALGGVFCRDARVAGVATGSTDPTARIIN ROLAEKPGAIALTIAETGQONAMI VDS TALPEQVIKDAVSATFSAGORCSALRYMYV QQDIABERVILLKGWALKVGPTEVRVSDGPVIDAEKAKGLEHIAALKAGKGLIA EKVPGNLNGHFVAVPAVEISIDELKENFGFVHLVRYAAQDLKVKVQAINATGFG LTMVHSRNEETARRIEELARVGNLYINRNQIGAVGVQVPGFGHGLSGTGPKAGGPNY LLRFVSRSTSYNTTAVGNSALSLADAE"
7020. .8540
/gene="putP"
/locus_tag="PA0783"
7020. .8540
/gene="putP"
/locus_tag="PA0783"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in *P. aeruginosa*); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 2"
/codon_start=1
/transl_table=11
/product="sodium/proline symporter PutP"


```

|||||
1811 GTTCCGAGAGCTTACACCGCGAGCTGTCCACCGCGGGGTGTACAGCTGACGCGG 1752
Qy GluHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeu 517
Db ATCTGGTGGTGGCGCGACCTTACAGTACACCGAGCGCGCGCTTCTTACGAGCTG 1692
Qy TyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsn 537
Db TAGCCACAGCTCCGCGCGCGCGCGTACCTACGAGGTAGCGGTCGCGACGCGGAC 1632
Qy LysGluArgSerAsnAsnLeuGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyr 557
Db AAGGAAAGAGGGGTCTCCACCGAGCTCCGCTCGCTTCGACACGCGGTGACACAGCGC 1572
Qy AsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTrpIleTyrAlaGlnThrLeuAsnAsp 577
Db AGCGTGGGGTGTCTACAGCGGCTTCTCCAACTACATC-----GGCTGCTCGCCAGC 1518
Qy GlyArgGlyProLys-----SerIleGluAspAspSerGluMetLys 591
Db GTCTGCCATCGCAACGAGAGGCGAAGTGTGCGCGCGCGGATGACGAGCGCTGCGG 1458
Qy LeuValAlaGlyTyrAsnGlnSerGlyValAspPheTyrGlyAlaGluGlyGlu----- 608
Db GAATACCTCTACAGCGGGTGTGCGCGGACTTACGCGGTGCGAGGCGGACCGGATC 1398
Qy ---IleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArg 627
Db CACCTGTCTGGAAGCGCGTACGGCACTTCGACCTGGAACCTCTCGGGGACTACACCGCG 1338
Qy GlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPro 647
Db GCCAAGAACAGGAC----- 1333
Qy PheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHis 667
Db ACCGCGCAACCGCTGCGCGGATCGCGCGCTGCGCTGGAACACCGCGG 1275
Qy LeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAla 697
Db CTGATCTGGAGATTG---CAGCAGTGGCGGCGCGGCTCGACGTCGACACGCGCGCTCG 1218
Qy GlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGly 707
Db CAGCAGCGGCTCCGAGGAGAACTCTCCACCGGAGCTAC-----ACCACCTCGCG 1164
Qy AlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrp-----TyrValLysAla 725
Db GCCAGCTCGGCTACAACTTCGACCTCGCGGAGAGCGCTGCTGCGCTTCTGTCAGGGC 1104
Qy AspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAsp---ThrPro 744
Db ACCACCTTACCAACAGACCGTGGCTAGCGCACTTGTATCTCGCGGACGCGGTGCGG 1044
Qy GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db GCGGCGGAGCGGATCGAGCGCGGGTGAAGGTGGCGTTC 1002

RESULT 23
AE016779/c 301995 bp DNA linear BCT 12-DEC-2002
LOCUS Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
DEFINITION AE016779 AE015451
ACCESSION AE016779.1 GI:26557023
VERSION
KEYWORDS
SOURCE Pseudomonas putida KT2440
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 301995)
AUTHORS Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D., Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,

```

```

Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P., Holtzapfel, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J., Timmis, K., Duesterhoft, A., Tummier, B. and Fraser, C.
Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301995)
Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D., Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R., Holtzapfel, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J., Timmis, K., Duesterhoft, A., Tummier, B. and Fraser, C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1..301995
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
89..712
/gene="sspA"
/locus tag="PP1320"
89..712
/gene="sspA"
/locus tag="PP1320"
/locus tag="PP1320"
/notes="similar to GP:9864180; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="stringent starvation protein A"
/protein_id="AAN66944.1"
/db_xref="GI:24982777"
/transl_table=11
/translation="MGATNRIACYSDDPADYSHRVLVLAEGSVSLDVPDAHLPKRIAEVNPYGSVPTLVDRDLALYESTVMEYLEERYPPMPYVPYVARGNSRLMHRIORDWCALATVLDPRSEAAETEARKEALRESLTGVSPLFSEFACFMSDQSLVDCCLLPILWRLPVLGIELPRQAKPLLDYMERQFAREPPQASLSVEREMRL"
693..1154
/gene="sspB"
/locus tag="PP1321"
693..1154
/gene="sspB"
/locus tag="PP1321"
/notes="similar to GB:X60221, SP:P24539, and PID:509291; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="stringent starvation protein B"
/protein_id="AAN66945.1"
/db_xref="GI:24982778"
/transl_table=11
/translation="MKCAFKEPLMNSRPLVRLVYEMVDNDCTPHMLVNAEPFKVQVDPGASDQIVLNTISPAVRSLSLNDNAVSFEGSPGVAHSLFVPGVAILGIARENGGWMFLEPSSNDGDELDVDDVQDDGPPDDGGQPPRPTGRPSLKVKV"
complement (1216..1794)
/locus tag="PP1322"
complement (1216..1794)
/locus tag="PP1322"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
/protein_id="AAN66946.1"
/db_xref="GI:24982779"
/transl_table=11
/translation="SKTPMRLGMLALTLCLSVTCSSVLTSRNSPTEDRGTRTIGSKIDPSLLETAAVNIKANPNLDKGSIVVSYNGIVLLAGOTPRADLKSLASQTASQVQRVKKVHNELOWQPSILARNNDAMLTKTKITQMLTDSAVPSRKIKVTENGIVYLLGLVTRQANSATAVVQGVGVQKIVKLFYIID"

```


QY 105 IleValValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 Db 143873 ATTCGATCTCGCAATGGCGCGTGCCTGGATGCTCGCTGCTTACGATCAT 143814
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleuArgGlyProValThr 144
 Db 143813 GCGTGGCGTGGACCGCGTAACTCGTCGAGCGGTGCAATGCTCGTGGCGCGCGCC 143754
 QY 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysPro 164
 Db 143753 CTGCTTTACGGCGCAACGCCATTTGGCGGGTGTCAACACCTTCGACACCGCATCCG 143694
 QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
 Db 143693 GACTCG---CCGATCGAGGCGCATCCAGCGTGCCTGAACTGCTTACGCGCGCGCGAC 143637
 QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
 Db 143636 ACCACTCGCAGCGCGCGCAAGCTGGAGCGCGCGCGCGCC---TTGCGCTGCGAC 143580
 QY 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys 224
 Db 143579 CTGGATGCCAACAGCGCGCGATTTCAACGACCTCGCATCCGCGCTATGACGACGCGC 143520
 QY 225 ArgLeu-----ProAspSerProArgPheAlaAsnGlyGlnHisArg 239
 Db 143519 AAGTGGCGCATGCCGAGCGCGCGCGCAACACCGCTCGAAAC-----143472
 QY 240 AlaValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
 Db 143471 -----AGCAGCGTCCCGCAGGAC 143454
 QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
 Db 143453 GCGGCGCAATCGGTGTGCTCTACCTGGAGCACCGCTACCGCGCTGCTGATAC---143397
 QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
 Db 143396 -----AGCCGCTAC-----143388
 QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
 Db 143387 -----GACAGCAACTACGCGTCC 143370
 QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
 Db 143369 GTGCGCGAGTCGGCGTGGC---CTGGACATGAACAGATCACTACGCTTCGCTCC 143313
 QY 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
 Db 143312 GAATCGCGCATCTCGACGCGCGCGTTCAGCTCGTCAAGGTAGATGCTGTATACCGAC 143253
 QY 360 TyrHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGln 379
 Db 143252 TACGAACACAGCAATCGAGGAGCGAGTTCATACCACTTCAGAACAGGCGTAC 143193
 QY 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
 Db 143192 GAAGCCCGCATCGAAGCCCGCTCAACAGCGCTCGGCGCGATTCAAGGGGTGATCGGTGC 143133
 QY 400 GlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnPro 419
 Db 143132 CAGTTCAGCGCGCAACGAATCTTCGCGCTGGC-----GAGGAAGCC 143091
 QY 420 MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlnGlnAlaAsnTrp 439
 Db 143090 TTGCTCCGCGCACCGATACCGACGCTGGCGGTGTTCATGCTGGAGCAG-----TGG 143037
 QY 440 -----AspAsnPheThrLeuGluGlyGlyValArgValGlnLysGlnLysAlaSer 456
 Db 143036 CAAGCCACCGCAACCGCTGAACCTTGAGCTGGCGCGCGCATGGAG-----CAC 142989
 QY 457 IleArgTyrAspLysAlaLeuIleAspArgLysAsnTyrTyrLysGlnProLeuProAsp 476

Db 142988 ACCGCGTAGACCGCGAGCCCAAGGCAACAGAACTTCGTCGAC-----142944
 QY 477 LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsn-----TrpTyr 494
 Db 142943 ---GCCGACGCGCGACGAGCTTCAACGCTTCAGCTGTGCTGGGGGTGTGTACGAG 142887
 QY 495 PheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThr 514
 Db 142886 CTCGACCCCAATCTGCTGCTGGCGCCAACTCCGCTGATACACGAGCGCGCCCTACCTTT 142827
 QY 515 GlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys 534
 Db 142826 TACGAGCTGTACGCAATATGGCGCGCATGACCGCGCGCTTGAAGTCGGTAGACGCC 142767
 QY 535 HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArg 554
 Db 142766 AGCTGTACAAAGAAAGGCCATTTCGCGCGCTGCGCTTCGACCAACGCGCAC 142707
 QY 555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIle---TyrAlaGln 573
 Db 142706 CACAGGCGCATGTGGTGTGTTCTACAGCCACTTTCGCAACTACATCGGCTGATGGC 142647
 QY 574 ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu-----589
 Db 142646 ACCGCAACACCGCGAGGCGGAGCATGACCATGAGGAGGAGGATCAGATCATGATCAT 142587
 QY 590 -----MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla 605
 Db 142586 GATCATGAGGCGTTCGCGGAATATACGATACCAAGCGCTGCGCGCGCTTCTACGCGATC 142527
 QY 606 GluGlyGluIleTyrPheLysProThrProArg-----TyrArgIleGlyVal 621
 Db 142526 GAGGCGCGAGGACCGCTGCGCATGTCGCGAGAACCGTTACGCGACGCTTCGCGTGAAGTTG 142467
 QY 622 SerGlyAspTyrValArgGlyArgLysAsnLeuProSerLeuProGlyArgGluAsp 641
 Db 142466 TCCGCTGACTATACCGGAGCC-----AAGAACCTCGACAGT-----142431
 QY 642 AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
 Db 142430 -----GGCGAGCCACTGCGCGCATCGCCCGC 142404
 QY 662 AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
 Db 142403 CTGCGCTGAGACAGCGCGCTGTCTGGGAGCTG---GACCGCTGCGACGCGCGGTAGAC 142347
 QY 682 TyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
 Db 142346 GTGACGACGCTGCTGCTGCGACGCGCAACCCGCAACAGAACCCAGACCGATGGCTAC 142287
 QY 702 HisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrp 721
 Db 142286 ACCACTTGGCGCGAGCGTGGCTTACCGTTTCGACATTTGGCGACGAGCGGTGGCTGGCG 142227
 QY 722 TyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerPheLeuSer 741
 Db 142226 TTTGTGCGCGCGGAAACCTGACCGACGCGCTGCTATGCGAGTTGCGATCTCGCTGCGC 142167
 QY 742 Asp---ThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 142166 GATATCGCACCGCGCGCGCGGTAGCTGGAAGTTCGGGTTCGCTACCACTTC 142113

RESULT 24
 LOCUS AE005695
 DEFINITION Caulobacter crescentus CB15 section 21 of 359 of the complete genome.
 ACCESSION AE005695
 VERSION AE005695.1
 KEYWORDS GI:13421330
 SOURCE Caulobacter crescentus CB15
 ORGANISM Caulobacter crescentus CB15

Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
Caulobacteraceae; Caulobacter.

1. (bases 1 to 11626)
Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,
Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R.,
Potočka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.,
Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H.,
Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J.,
Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and
Fraser, C.M.

Complete genome sequence of *Caulobacter crescentus*
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)

21173698

11259647

2. (bases 1 to 11626)

Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,
Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potočka, I.,
Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J.,
Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
Fraser, C.M.

Direct Submission

Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. .11626

/organism="Caulobacter crescentus CB15"

/mol_type="genomic DNA"

/strain="CB15"

/db_xref="taxon:190650"

97. .807

/gene="CC0206"

97. .807

/note="identified by match to PFAM protein family HMM

PF01144"

/codon_start=1

/transl_table=11

/product="3-oxoadipate CoA-transferase, alpha subunit"

/protein_id="AAK22193.1"

/db_xref="GI:13421331"

/translation="MVDVYKDAVEALAGLFDGNTIMAGGGLCGIPENLIAAREA

GTDLVVISNNGVDGFLGILLERQIKMVSIVGKLFQYLSCGELEFPNQ

GLAERIRAGGAGIPAFFRTGTGLVAREKTRFEDGEMVYMERGLTADLSIVKAWK

ADAGSLIYKRTARNPNMWTAGKTVVEELVAIGELDKDAHTFTGIYVDRLIKG

AKPEKIERVTTQKETA"

807. .1448

/gene="CC0207"

807. .1448

/note="identified by match to PFAM protein family HMM

PF01144"

/codon_start=1

/transl_table=11

/product="3-oxoadipate CoA-transferase, beta subunit"

/protein_id="AAK22194.1"

/db_xref="GI:13421332"

/translation="MAWTRDEVAAAKELQDFVYVNLGIGIPTLVANHIPGEMVTL

QSNGLMGMPFFYEEDADLINAGKQITLEDSSYFSSADSAMINGGHIASIL

GGNQVSENGDLANMVPGRKVGKMGMDLVAGVKRVVYVMDHCEKSGAPKLLKQCSL

PLTGAGVDLLITELGVFEINRGKTPVKLIELAPGTVDEVKATEAAFEAV"

complement (1449. .2438)

/gene="CC0208"

complement (1449. .2438)

/note="identified by match to protein family HMM"

/gene="CC0209"

/codon_start=1

/transl_table=11

/product="ThiJ/PfpI family protein"

/protein_id="AAK22195.1"

/db_xref="GI:13421333"

/translation="NFRFLVFLALLAATPGVAKPLVVLADQGVETTLIAPYAVL

ASSGVAVRVAPTIAPVRLMPQAWRPQTYAALTREPDVVIIPFLMPAYDPRDA

WIRAOQAGARIVSICAGETIARAGLIDGREATTHWELNRLAKAFKPKVRRQORW

IVDGPVTAGVTASIPASLALLRDLADGVNRETALNPLPFAAFVAGVRRVGTD

AAMSLGWLNAWRAEKVSVRIADGFDLLAFATLDAWSTRTNRSFAFASRSRHGL

TTIAPAKPPRAATTDPKDGDPDTKAKAIAAAYKATARYVALTITLHPYWLQO"

complement (2519. .4093)

/gene="CC0209"

complement (2519. .4093)

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK22196.1"

/db_xref="GI:13421334"

/translation="MQTADPASPPIENPAWRLTLMLMGGLTIVRLAALFLTPLELYP

DEAQYWLSELEAFYFSLKPMIAMLIMATTQIGDTEAWRLSAPFLWAGLALVHRI

ARLYCGWAGLAAALYSLMPGVLSGLINTADPLFLSLTWYSLPDAARRR

YALACMGALGLAFLSKTAAYVALGSHALHFAISSEARRRSPALVGLFVAFALVL

APNLWMAANQSTVKHTAANANWNAHOLFVRELIEFVSGQFVGVPVPAFLVGA

IMLVKRLQSPDLLLCPALPLIVAGEAFVSRANANAGAAVFGSVIVAGWLLR

WNAWRMLGLVLOAAPAFVACWPKVADAGLNGSKVRGWDQVTEAVIARVR

EEQALRGPLSAVAMDDRFVINAAYYGRDYFGQPGAPPLRMWHEAYFQNAETETPL

DADYGRALIVSLEGGYRPEIBODFKAVSGLQIARVRLDKTRSRVLDLFIAGFAFLP

RDPTGLPKPKTAAR"

4161. .6692

/gene="CC0210"

4161. .6692

/note="identified by match to protein family HMM"

/codon_start=1

/transl_table=11

/product="TonB-Dependent receptor"

/protein_id="AAK22197.1"

/db_xref="GI:13421335"

/translation="MWAGVAPPSPNLLNSINGRQETMRPNASVLMNEGFLQMNK

HPALATLALGALSASGAMAQSTGEVQVIVTASTVTRNQNCAIQAQSVPKARSI

TODFIRQIPQITLSDNMLPGVNFNDNDFAGSAGDINIRGDSORVALLDGIPL

NDSNGVYIYPNQMDSELIATKDVNQGTDDSPDTAAAGQTNVYTRKPYDFGVTV

EGQVGENFRFPAVDTGKVGWPSAFSYANAKNDIFNGFKTIYKQYNAIYQP

IGNGDPIFLIANYONRNFIIRMSLAQFNTGSGNTGFRDTCARPTGAGQIETNP

GNCKYKINPSNTGLRGOSLPHIENIITADPYQYTLANGGGRALPREDSGNTGS

CLVRGSALTGLDLNGDGLDVLVYGGENTVTRRYGVTSSLMKESQSVRAAYT

YDDAHKQTLIDNGDGLDVLVYGGENTVTRRYGVTSSLMKESQSVRAAYT

AYRGRPMEDVLVNVGRAPFPKRDLLNNFCYQADTFNAYCQADPAVAGSDGAGQ

TLVTLPGQGTQYGRPKRQYDDVLNVGVSLEYAKQYTVISVAETISVPTDOL

YDKLITPEPTQAYDLGYRQSPITLFAASLYTNKQRIERAFNEQDNIASFNV

GDVYKGFDAQLGKPEQLSFYASFIIDTELKANIPTGLTSLVPTKGRVYETPK

YQGGVGNWDTIRALSLGVNAKIVGERWNLNTKAPGSLVNMVRYELDGLGLKD

TYLQINARNLFNERYLGLAVNPSGTGLGQPGAPRTFVATIHARF"

6747. .7859

/gene="CC0211"

6747. .7859

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK22198.1"

/db_xref="GI:13421336"

/translation="MSPSFAKEARLDQGVVVRGHLRPLRFLSLTAFHWACRMVTVPAEW

APHRMVGFPFSAHLQWEDLEQAOQEVADLALAGPAGARVLRVMDVGEAARARA

LLSDTVEIVRGQDGLMRLDTGPIFVDDAGKAAVAGFKENGWGVKSLGGDDIVAEQ

IAASGAPLVNDFILEGALDHGNGTILTTQCLLNDNRNPGMBEATNAAALTEAL

GAKYVLMWGLKNDHDTGHVDNLRFAVACPMGVLGDDFNADITGDTAKMLAG

MTDSRPSQVVRIPSPGKLLDEGFPFASHMFLIANEAVIPIYASESGAFAVEV

IKGLFPERQVIGLPSFTALTGGSGSFHCISQOEPETA"

7859. .8737

/gene="CC0212"

7859. .8737

/gene="CC0212"

/gene="CC0212"

```

/note="identified by match to PFAM protein family HMM
PF00795"
/codon_start=1
/transl_table=11
/product="hydrolase, carbon-nitrogen family"
/protein_id="AAK22199.1"
/db_xref="GI:13421337"
/transl_table="MARTLSVAATQTSYGMQLQNIKKTEGFIREAAKGAQVILPSE
LFQGPYFCVAEEBWFQAHPREHPVVKAIAPLAGLGVVPIPSIPERGPHVFNLS
VMADGSLMGVYKSHIPDGPYMEKYFPGDTGPKVWDTRFGRIGVIGICWDQWYP
ECARALMGALFYTAIGSEPHDASLDTALPWRAMQGHAVSNVIPVIGANRIGF
EPWDYNGGQGTFTGSSFSFVADHRGDLVSELGRADEGLVSATFDLDTTHRAAWGFR
DRRDELYTALSSGRPA"
/gene="9282"
/contig="CC0213"
/contig_start=1
/contig_end="9282"
/contig_size="CC0213"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="acetyltransferase, GNAT family"
/protein_id="AAK22200.1"
/db_xref="GI:13421338"
/transl_table="MIETDLVLRPAVDADATADLNARPEVGAWLGVSRAESDA
FVDMARHAQNGGFWVERKADVLVGMGTGVMVPPAMNMDPTTEIGHRFLEPVWG
EGDPLRATLALVGFVLRPELIAFTARSNLASQAVMRIGMTAPAPDFLHPLRA
EGDPLRATLALVGFVLRPELIAFTARSNLASQAVMRIGMTAPAPDFLHPLRA
9410..11473
/gene="CC0214"
/contig="CC0214"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAK22201.1"
/db_xref="GI:13421339"
/transl_table="MVSGPMIRVVALSAALVLAATAAHAKDAAPASQADAEV
SKVIVTAAPGVSDALATASAVLDRTDLDAAPKGLDALAGLPGVRSFTFGAGAR
FVVRGLAGFVOVLITNGVQIDASALSPDHQATDPGEAREIEVLRGPAALAYGSAI
GVMNIIIDR1STQPDIGMSRLLAGTDDSYALSGAVHATVGMVLTDLAKRE
SKDYKIPVPSARQLAEGETAGGAGRLNSAVLETFGAGLSYVVDKGFQWMSIK
RTDSTYGVPHGHEGEAEAGHEESAVTIGLKOTRIDLRGEVDALDGRFAKVRFSG
GHADYTHTEFFSGAVGTFTSDGYEGRELVQTERGKWKAGVQALRRNFDAIGDRA
YVPTKITEFGAFTQLELDGYEGGLRDTRELDLKGKADFTWISSVGAWRP
TTSEFVGLALSRSSRTESELFAGPHATRGFEIGDADLKEETATSLREATLHYGE

```

Alignment Scores:

Pred. No.:	1.13e-28	Length:	11626
Score:	636.00	Matches:	206
Percent Similarity:	40.23%	Conservative:	107
Best Local Similarity:	26.48%	Mismatches:	297
Query Match:	15.76%	Indels:	168
DB:	1	Gaps:	23

US-09-936-377-2 (1-758) x AE005695 (1-11626)

```

QY 22 AlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys 41
Db GCTTCGGCGGCAAGCCGCGAGTCTCGAAGGTCTATCGTCCCGCGCACCTACGGGTC 9577
QY 42 SerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle 61
Db TCG-----GCTGATGCGCTCACCGCCAGCGTCCGCGTGTGCTG 9613
QY 62 SerGlyAspThrLeuArgGlnLysAla---ValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db GATCGTACCGACTGATCTGGCGCCCGCCCAAGGGCGCTGGCGCGACCCCTGGCGGCTG 9673
QY 81 ProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db -----ACCAAGATCACCGAGTTCGGCGCGGCGCCAGCGCGTGTGCGGCTG 9733
QY 440 AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyr 459

```

```

QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db GCGGGCGCGGTCTCCAGTCTGACCAACGCGTGGCGCAGATCGACCCAGCGCGCTG 9793
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLysLeuArg 140
Db TCGCCCGACGACCGAGTCCCGACCGATCCCGGAGCGGAGCGGATCGAAGTGTGCGC 9853
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAlaAsp 160
Db GGTCCCGCGCGTGGCTTATGGCGCTCGCGATCGCGGCGTGGCTCAACATCATGAT 9913
QY 161 GlyLysIleProGlyLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
Db GATCGGATCTCCACCCAGCAGCGATCGACGCGCATGAGCGGTGCGCTGCTCGCGCG 9973
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db GGGACCGGTGACGACAGCTACGCGGTGTCGGCGCGCTCCACGCCACGCTCGGCGCG--- 10030
QY 201 PheValLeuHisThrGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db ATGCTCTCGAGCTCGACGCGCTGAAAGCGCGAGAGCAAGGACTACAAGATCCCGGTCTAT 10090
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAla----- 234
Db -----CCGGAGTCCGCGCGTCACTGTGCGCTCGAGGCGCGAGCGGCC 10132
QY 235 ---AsnGlyGlnHisArg-----AlaVal 241
Db GAAGTGGCGCGGTGCGCTTGAGAACAGCGCGTGCATCTGGAACGTTGCGCGCGAGGC 10192
QY 242 LeuGlyTrp-----ArgLysArgPheTyrArgThrTyrSerAspArgArgAspGln 259
Db CTCTCTCTATGTCGCGCACCAAGGCTTTGTGCGCATGTCGATCAAGCGCGACGACTCCACC 10252
QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
Db TATGGCGCTCCGCGCGCACGCCCATGAG----- 10279
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
Db ----- 10279
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
Db -----CATGAGCGCGCGAGCGCGCGCGG 10303
QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db CACGAGGAGAAAGTGGGTCCACCTCGGTCTGAGCAGCAGCGGTATTGACCTGCGTGGC 10363
QY 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db GAGTACGACGCGCATCTGGGCGCGCTTCGCAAGGCTGCTTCGCGCGCGCGCGCGGAC 10423
QY 360 TyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGln 379
Db TACCCACACCGAGTTTGGGTGAGCGGTGCGCGCATCGGACCCAAAGTTCACCTCGGATGCGCTAT 10483
QY 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
Db GAGGGCGCGCTCGAGCTGTTGAGCGCGGTGCGCGGTGGAAGCGCGTGGCGCGTCC 10543
QY 400 GlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnPro 419
Db CAGGCTCTACCGCGCAACTTCGACGCGCATCGCGGCGAGGCGCTATGTGCGCAAG----- 10597
QY 420 MetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp 439
Db -----ACCAAGATCACCGAGTTCGGCGCGGCTTACCCAGCAACGCCCTCGAGCTT 10645
QY 440 AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyr 459

```



```

NCVTGAG"
2252..3016
/gene="blr2224"
CDS
2252..3016
/note="ORF ID:blr2224"
/codon_start=1
/transl_table=11
/product="ABC transporter amino acid-binding protein"
/protein_id="BAC47489.1"
/db_xref="GI:27350479"
/translation="MALKRFVQAAVAALAIAGVVPASAQOLKVGSTPTGTFELDT
KNTIIGIMWDLITEGKDLGNVQIEPMSFSALESLSSKIDIIIAAMFITAAPRKE
VWFSPIFYTGSLVVPKNDTKAYATODDKGETVGAQVGFADALKSGLEADVK
AVDTIPILRDVNTGRLKAGYADYPILAVNLKOGFPEVRLVDSYKPVTVSGVIGVR
KGSALLGKINASLAKLAKNGTTDKLEKWKQRAQ"
3045..3698
/gene="blr2225"
3045..3698
/gene="blr2225"
/note="ORF ID:blr2225"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter permease protein"
/protein_id="BAC47490.1"
/db_xref="GI:27350480"
/translation="MKGFWDHAAEFPIILMSGVALTIVVTIGSLLSVLGIWAMMR
VSGIKVLSMLSGIPIVIGIPIVILLFYLIFVMPDLGVTLSALQAAIIGLAIYSAY
QAENFRAGIADIKQIEAAQSIGMGWLTMRVLPQAVRIVLPPYGNVMIMLKDS
SOASTIVASLALQGLIASSTFKNTSVFTLVALMTWSIPLILLVRHFEKRAKK"
3695..4426
/gene="blr2226"
3695..4426
/gene="blr2226"
/note="ORF ID:blr2226"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter ATP-binding protein"
/protein_id="BAC47491.1"
/db_xref="GI:27350481"
/translation="MIELTDVHSFGKVEVLGITKVEKGEVVCVPGSGSKSTIL
RCINGLESYGEINVEGLKVDKDPASIVNVRTQVSMVFORFNLPHRTVLNVNVEGP
LYNKEARQVLEGRGALLAQVGLAEKADAPPLQSGGQORVAIARALAMQKALIF
DEPTSAIDPELVGLVGRKLADGDMTWVVTHEMGFARDVADRLVIDGGVIVFQGG
PAKSVLQROHPRQDFLRVLHPL"
4439..5725
/gene="blr2227"
4439..5725
/gene="blr2227"
/note="ORF ID:blr2227"
/probable oxidoreductase"
/codon_start=1
/transl_table=11
/protein_id="BAC47492.1"
/db_xref="GI:27350482"
/translation="MTARLPLPSSIYADTAVAPVATPPLDKTVSAIIGGYTGLS
TALHARQGEALVLEAQEGCWGASGNNGGTNPGLKHPDQIEADFGTGLGRMTIEF
SYGPNFTHELIRYQIIPCBARQNGTLRAAYNEASAAAIKTAQCICIRRMGPVYTLNR
EOLREMTGDRYICAMLDARGDDLPLSYARGLARAI SAGAKVHGETPALSLRDEGA
HWLETRAVVRAEKVLLATNGTDDLPALRSTIVPTVSSIVASAPLSEVARSIMP
TRPVYSGHTIVYRIDQNNRLMGGEGPMWISSPDSVAYLIRYAEIRLWPLQKGA
WTHGWNRLAITKDYHPHVEPAESILISLGCNGRGVALSTAMGAQLARLLIGGAKAE
IDMPTVTKIPIMHAFWPEVGVTTAVITGRVDRRLGI"
/gene="blr2228"
3615..6145
/gene="blr2228"
/note="ORF ID:blr2228"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC47493.1"
/db_xref="GI:27350483"

```

```

/translation="MCQHALSRCAIHLRERHDSSTFAGNLAHSPFIDRTSALAPLE
NMKRLAASLIIAAGFTLTNPAPAFANGHGCHGYGHGYGHAMGHVYDGTTPQWSHG
RKVMRGHGCPPEGLWKQRC"
complement (6201..7916)
/gene="blr2229"
complement (6201..7916)
/gene="blr2229"
/note="ORF ID:blr2229"
similar to ACY1 bifunctional acylase"
/codon_start=1
/transl_table=11
/protein_id="BAC47494.1"
/db_xref="GI:27350484"
/translation="MFSLAAMCHCPRASPNALFEKSPSQVHRSQDORQOTQRIIMP
HOFSLAQTVRPVAVTKGIVAAQSRRAAEVGAQVLAAGGCDVDAIVATTMALNVLEP
WNSGIGGGGANVLYARENRTEVIDYGMCAPOSRLRVADYPLSGGGAADLPFPVRVKD
DRNTHGCAIARVGVAGVAGHRAHAKLPKDIAPAALAGSLADMMWTTLTIAS
SAUDLRYPASAAAFKDLGFLPTAPWGIKSTRLSQKLIKATLSHLAEAGRDFYQGD
LAKSIASDIKADGSSLSVEDIAAFRAHLREPLAIPYRGKVFATPELTAGPTLAHLR
LLQNLKPAGAPDAATYAEYALALQAAAYRERLKDMDGADGKRSIGASYLAPACTHFS
VDRHGNVAAVTQTLISFGSGKYVTPHTGIMNNGIMWFDPTGTTSLAPFKRCLTN
YTPVIAETRDGRLAIGASGRRILPAVMQLASFAMDFGMDLDAIHQPRIDASEGAV
VIGDVLDPADTRKALAAAFDEESRVQALPQKFACPSPVVMREGEVNFQAVEIFQPWD
AVAEA"
7934..8707
/gene="blr2230"
7934..8707
/gene="blr2230"
/note="ORF ID:blr2230"
/codon_start=1
/transl_table=11
/product="enoyl CoA hydratase"
/protein_id="BAC47495.1"
/db_xref="GI:27350485"
/translation="MAGVKQARDGAVGILTDEPASINAMTPDLIGALAAAAGVGMQD
EGVRALILTGAAGPFCGONLKASEALGEDIAAGVMRPFYPAFKALRECRYPVVAVN
GVAAGGSLAMAGDIIVAARSASFIQVFSIALVDPDLSGTWLLPRLNGRORALEML
LINEPTARAOEIGLIRVQVDDKLMGEALVLRARLAGPTALVATRALVESEHAT
YEAQFRRIEIQMLIRKSDALEGRDAFVEKKAFTGR"
8664..9227
/gene="blr2231"
8664..9227
/gene="blr2231"
/note="ORF ID:blr2231"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC47496.1"
/db_xref="GI:27350486"
/translation="MPSSRSARRSPGGEPTKSMPTGCPGIASLENRFAASGETTAS"

```

Alignment Scores:

Pred. No.:	9,09e-17	Length:	300400
Score:	469.00	Matches:	200
Percent Similarity:	38.4%	Conservative:	124
Best Local Similarity:	23.7%	Mismatches:	335
Query Match:	11.6%	Indels:	184
DB:	1	Gaps:	34

US-09-936-377-2 (1-758) x AP005943 (1-300400)

```

Qy 4 ThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGln 23
Db 283418 ACAGCGCGCGAGCCGATGTCGCGAGGTGGTACCGACCGCGCGCGTCCA-TGTGCG 283360
Qy 24 AlaHisGluThr-----Glu-GlnSerValGlyLeuGluThrValThrValGlyIly 41
Db 283359 GCGCACCGGACGCGTCCGCGAGCCACACACGCGCGCGAGCTACGCGCGCGCG 283300
Qy 41 sSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAs 58
Db 283299 GTCTGCCCGCGCGCTCAGCAGGCGTGTCTGCCCATCGTGACCAACACGAGTTCGCCCGCT 283240

```


QY 58 pYlaileSerGlyAspThrLeuArgGlnLys--AlaValAsnLeuGlyAspAlaLe 77
 Db 283239 CACGTTGGTCCGACAGAGATCCCGCGAGGGCGGCGTTCAGCTCGCGCATCTCTT 283180
 QY 77 uAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAla---ProVa 96
 Db 283179 GTTCTCCAAAGCGCGGCATCACCGCTCGAGCTTCGCGCGCGCGCTCCAGCGCGCGAT 283120
 QY 96 lIleArgGlyGlnThrGlyArgGlyIleValLeuAsnHisHisGlyGlnThrGlyAs 116
 Db 283119 CATCCGGGGCTCGACGTCAACCGCTCGCATCGTCGAGNACCGCACCAATTCGGGTGG 283060
 QY 116 pMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnInVa 136
 Db 283059 CGCTCCGACCTCGCGAGGATCACTTCGTCGCGATCGCTCGCGACCAACACAGGT 283000
 QY 136 lGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVa 156
 Db 282999 CGAGTGTGCGCGCGCGCTGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 282940
 QY 156 lAspValAlaAspGlyIleProGlyMetProGluAsnGlyValSerGlyGlu-- 175
 Db 282939 CAGCGCGACCAACACCGGATTCGAGTTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 282880
 QY 176 -LeuGlyLeu----- 178
 Db 282879 CTACGGCTTCGCGACCAAGGCGCGCTCGCGAACCGCGAGACCTCGCTCGCTCGCG 282820
 QY 179 -----ArgLeuSerSerGlyAsnLeuGlyLeuThrSerGlyGlyIleAsnIleGl 196
 Db 282819 CGAGACGCGCTCGCATTCAGCTCGCTCGATCGCGCGCTCGAAAGCGCGTGTCTCGA 282760
 QY 196 yLeu---GlyAsnAspPheValLeuHisThrGluGlyLeuTyrArgIysSerGlyAspTy 215
 Db 282759 CACCGCGCGCGCAATTTCCTTCATCGCGAGCTCTATGCGCGCACCAACCGGACTA 282700
 QY 215 rAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgPheAlaAs 235
 Db 282699 CGGCATCCGAGCTATCTCTGAAC-----GACCAGACGGATCCGTCGCCAA 282649
 QY 235 nGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAs 255
 Db 282648 CGGC-----CGCCACCGCAACTCGGC 282628
 QY 255 pArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAs 275
 Db 282627 GACGCGTCCGACGGCGCTCGATCGCGGCTCTCTAC----- 282591
 QY 275 pIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLe 295
 Db 282590 ----TTCCTCCAGGGCGGTATATCGGGCT-----GCGAT 282559
 QY 295 uThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAs 315
 Db 282558 CACGAGACGACTCGCTCTACCATCCCGGATC-----GACGGCGC 282514
 QY 315 pAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTy 335
 Db 282513 CGAC-----CACACACGCGC-----ATCGAGGTCCACAGACCAAGAT 282475
 QY 335 rGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLe 355
 Db 282474 CAACTGCAAGGGCGAGTACCGTCCCGATGCGCGCGGATCGACACGATCCGGCTTCGGGC 282415
 QY 355 uAsnArgAsnAspTyrHisHisAspGlu-----LysAlaGl 367
 Db 282414 CGCGCGACCGACTACCGGCACACAGATCGGCTTCGCGATCCCGCGACCCCAACAG 282355
 QY 367 yAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHi 387
 Db 282354 CGACGCGGTGCGCGACACCTTTACCAACAAGAGAGAGAGATCCCGACCGAGGTGACCT 282295
 QY 387 sGlnPro-----IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGl 403
 Db 282294 GATCGCGTTCACACGCGCGCTTCGCGAGGTGACGACGCGGTGGGCTTCACGCTCGGCA 282235
 QY 403 yGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAs 423
 Db 282234 CCAGAACTGACCGCGCGAGCCCGGACCAATCCCGGACGCTGTTCAACGCGCTTCGGA 282175
 QY 423 p-----AsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTr 439
 Db 282174 CCCCACACACACGCGGTTCGCTTACGCTTC----- 282138
 QY 439 pAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTy 459
 Db 282137 -AACGAGTTCAAGTTCAACGACCGCGCGCGCGAG---ATCGCGCGCGCATCGACA 282082
 QY 459 rAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAl 479
 Db 282081 TGTGAGTGTGACGCGCACGACCGCCGAAATTCGCGCGGACATTCCTGCGCGGAC----- 282030
 QY 479 aHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHi 499
 Db 282029 -----GGACGCGCGACGCTGCGCATCGCGCAATCGCTCTACACGCGCAAGAG 281980
 QY 499 sLysLeu-----SerLeuThrAlaSe 506
 Db 281979 CGGCAGCATCGCGCTGTTCACGACCTCGCGCGCGCATGTCGCGACGATCACCGCGCA 281920
 QY 506 rHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaTh 526
 Db 281919 ATATGTCGACG 281860
 QY 526 rAsnThrPheGluValGlyAsnLysHisIleAsnLysGluArgSerAsnAsnIleGluLe 546
 Db 281859 CGCGAGTTCGACATCGGCAATCCGAACTGACGATCGAGACCGCAAGTTCGTGAGCT 281800
 QY 546 uAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPh 566
 Db 281799 CGCGCTGCGCAAGGCG 281740
 QY 566 eGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAs 586
 Db 281739 CGCAATTCATCTTCGCGCGCTCACCGCGCTGATGTCGCGC-----GA 281695
 QY 586 pAsp-----SerGluMetLysLeuValArgTyrAsnGlnSe 598
 Db 281694 CGACTTCGCTCGTGGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 281635
 QY 598 rGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrAr 618
 Db 281634 GAACGCGAACTTCGTTGGCGGCGGAAATTCAGTCGCGAGCTC---GACGTGCGCGCGTCCA 281578
 QY 618 gIleGlyValSerGly-----AspTyrValArgGlyArgLeuLysAsnLe 633
 Db 281577 GCGCGCATCTGGGGATCGAGAACCGCTCGACGTGTTGCGCGCACCTTC----- 281526
 QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGluAspAs 653
 Db 281525 -----ACGACCG 281518
 QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
 Db 281517 CACCAAGTTCGCGGATTCGCGCGCTCAAGATTCGCGCGTCAAGATTCGCGCGCTGT 281470
 QY 673 rAspArgIleAsp-----AlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLys 690
 Db 281469 CTGCGCGACGACCAACTGCGTGTCAACCTGTTGACGCGCTTCGCGCAGAACAA 281410
 QY 690 sLeuAlaArgTyr---GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAs 709
 Db 281409 CGTCGCGGTGATCGCGAGACCGCGCGAGCTACACCTCTCTGAGCGCGCGAGTGGAG 281350
 QY 709 nTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyr--- 722


```

/translation="WERSRSIFFIMLRKTNFQGLNLSYIDTNSKSKPTILLCHANGY
SAFTYKVESLONSYSVIALDPAGHGESDSTLNFDRWYFRDQVLSLISENNVNI
GIGSLGASLLASYSRDPKEXKVIADHPDPTLNFLOITYSRIFNPLAQVAIKRRE
FKNDTVKRIYKTRPSRWDGDIYDDIQSCFRLGVNGEAILKCAPEVBAKIFDSVS
YRSIFQFKIKITETHITIPDPHEVCSPNTAKKIISGNRSLSLEIWPKTSHFFPFEEKH
KTLDRILEVLD"
6714..7046
/locus_tag="LA3148"
6714..7046
/locus_tag="LA3148"
/locus_tag="LA3148"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:24197045"
/translation="WILQIRKSGFLSILLIIFVSSGLHIIHVEQEIQKFSGYSTNKD
TKQIVACICQFQRESNLNPNFLTNQPSVFLKIRISFVIFIHFNYPQRFGR
PPFLUPLS"
7092..9425
/locus_tag="LA3149"
7092..9425
/locus_tag="LA3149"
/locus_tag="LA3149"
/codon_start=1
/transl_table=11
/product="hemin receptor"
/db_xref="GI:24197046"
/translation="MKKQYFYFRIFLSIIPISILALDVTITGIVKNGNSIN
VKIIVQESRKTTITNEKGFVNLHVPKGYTLVAISRGYQSETISITIGKDEKTOFT
LLESIDOSAINVTAKTNSDPLTAPOITVLSGRLDRORGETAMGAINTPGVSNL
TGSGTSPLRLGLTQORVLAWTDGIRQEOQFGDDHVELDSFNIQKIEIIRPGSIL
LYGSALGVVNVIRDKAPLSGEGIPKAGIPNSNSVNNKQDAGNPAIYGNLNGFY
RASNSRAGKITTNGTRNTGMIERKQNSASIGLDGMGNFYLDSPFREOTQDLON
PNEPFGATVSKLHEKSHSFFISAGNLELDFSYORNNREIESKNKLLPDIOL
LDENAEVLDSQFTQVTSKEXIYGLNLFLLDTTADAKFKPKIFNFKLNGTIGSLE
OMKTIETGPIIPISIIAGYFLEELKGLSFTSAGIRGDKRSVDLRNNOTIGSFE
CTKMYMTGLSGVNVIVKGFSTIFNYGRGFRAPTPPELFSYVHEGTGKFEYGNH
LKPEYNLDSLRVASKIOETISVFNHQNFIYAASIAQIDLDSGLPKYEVKQGN
ALIKGESIQAEIRKLVFSGGDIIVSRMGSDTHPLPTEPRABAGLAWTDSIL
GLNFIYSINGFYDSQVRDPKETPTKGYMLIDVGLGFELPHFGDGTSGKPTVDLSIQ
NVFNYSYVDHLSRYDYALNPLGNAILKVSFPFFTAIP"
9436..9810
/locus_tag="LA3150"
9436..9810
/locus_tag="LA3150"
/locus_tag="LA3150"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:24197047"
/translation="MKKYLIFLTLILFCQNEKNNQDLMLRLLANTFNETRSDCV
YCTDQAFQNGCSFSPNIPWSCQGYSSGRQKNSIRISCEDLTSKGWFPENSNOGA
KSCSPSLSCPPEAYRAFTPDGF"
complement(9855..10016)
/locus_tag="LA3151"
complement(9855..10016)
/locus_tag="LA3151"
/locus_tag="LA3151"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:24197048"
/translation="WNSYFKIFPSRFLTNSRYLELLKNFIAVNKNKIASIYFFHTEI
DVEFTFOKL"

```

BASE COUNT 3177 a 1970 c 1645 g 3422 t

Alignment Scores:

Pred. No.: 3,286-14 Length: 10214
 Score: 396.00 Matches: 174
 Percent Similarity: 39.95% Conservative: 130
 Best Local Similarity: 22.86% Mismatches: 303
 Query Match: 9.81% Indels: 155

```

DB: 1 Gaps: 29
US-09-936-377-2 (1-758) x AB011477 (1-10214)
Qy 28 GluGlnSerValGlyLeuGluThrValThrValValValGlySerArgProArgAlaThr 47
Db 7404 GAAAGTTCTTGGACCAATCCGCGATCAACGTCACCGCTAAATCC-----ACA 7451
Qy 48 SerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu--- 66
Db 7452 AATTCAGATTTTAACTGCTCCACCAACCAATACCGTTTTCGGTAGACAAATAGAC 7511
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db 7512 CGTCAAGGGGAGAAATCGCAATGTCTGTCATCAACAATCTCCGGGAGTTTCCACCTT 7571
Qy 87 GlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgGlyLeuLys 106
Db 7572 ACAACAGGATCAGAACTTCTAAACCGATCTCCGAGGCTTACCAGCAAAAGGGTTTA 7631
Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
Db 7632 GTTATGACC---GACGGTATTGCTCAAGAAGAACAACAATTTGGTCACGACCACAGTA 7688
Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu 146
Db 7689 GAATTAGATCTTTCAATATTCAAAAATTTGAAATATTAAAGAGCGCCGGGTAGTTTACTC 7748
Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro----- 164
Db 7749 TAGGGCTCAGACCTCTTGGAGGTGTAGTAATGTCATTCGAGACAAAGCACCTTTTCC 7808
Qy 165 ---GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly 183
Db 7809 GGAAGAAGAAATCCAAAG-----ATGCGAGGCAATTTTCAATTCCAACAGTATTCTAAC 7862
Qy 184 AsnLeuGluLysLeuThrSerGlyGlyIleAsnIle---GlyLeuGlyLysAsnPheVal 202
Db 7863 AAC-----AAACAGATGCTGCACTTGTCAATCTACGAAA--TCTAAACGGTTTCGG 7915
Qy 203 LeuHisThrGluGlyLeuTyrArgLysSerGly---AspTyrAlaValProArg----- 219
Db 7916 TTATCGGTGTAGCTCCAATTCAGAAAGCGGGTAAATTTACTACTCCGAACGGAACCTT 7975
Qy 220 -----TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGln 237
Db 7976 TCGTAATACAGGATGATAGAAAAAATCAAAATGCGTCCATTCGTTGACCGGGGATG 8035
Qy 238 HisArgAlaValLeuGlyTrp-----ArgLysArgPheTyrArgArgThr 252
Db 8036 GGGTAATTTTATTTGATTTCTTTTAAAGGGAACAACCAAGATCTTTTACAGAA--- 8092
Qy 253 TyrSerAspArgAsp---GlnTyrGlyLeuProAlaHisSer-HisGluTyrAspAs 271
Db 8093 ---TCCAACGAAATCTCTGAGCTACTGCTCTCAAAAACATTTCTTCCAGAGAAATCTCA 8149
Qy 271 PcyHisAla-----AspIleIleTrpGlnLys 280
Db 8150 TTTTCATCTCGTTTATTTTTCGCGAGAAACTTGGAACTTGATTTTCTCTATCAAG 8209
Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
Db 8210 AAAC-----AATCGAAGAGAAATCGAATCTAAAAATAAATCTTCAATCAAGATAT 8263
Qy 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
Db 8264 TCTTTTG-----GACGAAATGCGAAGT--- 8288
Qy 320 sAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGl 340
Db 8289 -----TTAGATAAATCTTTTTCAGTTCTATCAAGTACTTCCAA 8326
Qy 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360

```

```

Db      8327  AGAATAA-----TACCAAGGTTTAAACCTTTTCTAGAC----- 8360
Qy      360  rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhPheAsnAsnGlnThrGlnAs 380
Db      8361  -----TACCAAGGTTTAAACCTTTTCTAGAC----- 8368
Qy      380  nAlaAArgLleGluLeuAArgHisGlnProIle---GlyArgLeuLysGlySerTrpGlyVa 399
Db      8369  AGCGAGCGTAAATTCACCATAACTATATCAATTTTTCAGAAAGGAACCTATCGGAAT 8428
Qy      399  lGlnTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnPr 419
Db      8429  ATCTGGCCTGGAACAAAGAAATAAACAATCGGTACA-----GAGCC 8470
Qy      419  oMetLeuAspAsnLysValGlnHisTrpSerPhePheGlyValGluGlnAlaAsnTr 439
Db      8471  GTTGATTCCTCTACGCTATTATTAACATTCGCGGTATTTTATAGAGAATCTAAACT 8530
Qy      439  pAspAsnPhThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTy 459
Db      8531  GGGTTCCTTAACTTTTCAGCGCTGGAATTAGAGGTGATAAACCGTTCGTCGACATTCGAAA 8590
Qy      459  rAspLysAlaLeuIleAspArgGluAsnTrpTyLysGlnProLeuProAspLeuGlyAl 479
Db      8591  TAACCAACC-----CTTGCTGT 8608
Qy      479  aHisAArgGlnThrAlaAArgSerPheAlaLeuSerGlyAsn-----TrpTrpPh 495
Db      8609  CTTTCGAACAAACCAAAATTTATATATGACACCGGTTCACCGGTGTTGTAGGAGAT 8668
Qy      495  eThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGl 515
Db      8669  CTATAATCTTTTCCACCATTTTATTAATACGACGCGGTTTAGGCGTCTACTCCTTT 8728
Qy      515  nGluLeuTrpAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHi 535
Db      8729  TGAGCTTTTCTCATCGAGTCCACGAAGAACCGGTAAATTTTGAATTTGGAACAATCA 8788
Qy      535  sLeuAsnLysGluAArgSerAsnLysLeuGluLeuAlaLeuGlyTrpGluGlyAspArgTr 555
Db      8789  TCTAAACCTGAAATTAATCAATATCTGGATTTTCTCTAGATTTGCTTCTTAAAT 8848
Qy      555  pGlnTrpAsnLeuAlaLeuTrpArgAsnAArgPheGlyAsnTrpIleTrpAlaGlnThrLe 575
Db      8849  TCAAACTGAGATAGCTGTTTTCAAATCATATTCAAATTTTATATATGCTCAAGTAT 8908
Qy      575  uAsnAspGlyAArgGlyProLysSerLleGluAspAspSerGluMetLysLeuValArgTy 595
Db      8909  C-----GCCCAATAGACTTGGATTTCTGTTTACCAAAATACGAGTA 8950
Qy      595  rAsnGlnSerGlyAlaAspPheTrpGlyAlaGlu-----GlyGluIleTrpPh 611
Db      8951  TAAACAGGAAACCGGATTTTAAAGAGGAGATTTTCGATTCACGACGACTTTCG 9010
Qy      611  eLysProThrProArgTrpArgLleGlyValSerGlyAspTrpValArgGlyArgLeuLy 631
Db      9011  TAAACTCGTT-----TTTTCGCGTGAATCGATTTGTTTATCTAGAAATCAA----- 9059
Qy      631  sAsnLeuProSerLeuProGlyArgGluAspAlaTrpGlyAsnAArgProPheIleAlaGl 651
Db      9060  -----AG 9061
Qy      651  nAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLysLysAlaSe 671
Db      9062  CGATCTACTCTCTCTAGAACTACACCAAAATCGAGACGCGGTCTCTAGATGGAC 9121
Qy      671  rLeuThrAspArgLleAspAlaAsnLeuAspTrpTrpArgVal-----Ph 686
Db      9122  GGAAGATTCTATTTAGGTTTAAAACTTCTACTTTTCCATCAACGGTAGGTTTATGA 9181
Qy      686  eAlaGlnAsnLysLeuAlaArgTyGluThrArgThrProGlyHisHisMetLeuAsnLe 706

```

```

Db      9182  CTCTCAATATAGAGTCGATCTCTAAAGAAACACCTTACCAAGGTTTATATCTTATAGACGT 9241
Qy      706  uGlyAlaAsnTrpArgAsnThrArgTrpGlyGluTrpAsnTrpTrpValLysAlaAs 726
Db      9242  TGGACTCGGTTTGAACCTT---CCTCATTTCCGAGACGGAACCTTCAAGGCTACTGTAGA 9298
Qy      726  p-----AsnLeuLeuAsnGlnSerValTrpAlaHisSerSerPheLeuSerAs 742
Db      9299  CTTGAGTATTCAAAATGATTAATCAATGTTTCTTATGATGATCAATCTCAGTAGATATAAGGA 9358
Qy      742  p 742
Db      9359  C 9359

RESULT 27
LOCUS   AE016931/c
DEFINITION Bacteroides thetaiotaomicron VPI-5482, section 6 of 21 of the complete genome.
ACCESSION AE016931 AE015928
VERSION   AE016931.1 GI:29338516
KEYWORDS
SOURCE    Bacteroides thetaiotaomicron VPI-5482
ORGANISM  Bacteroides thetaiotaomicron VPI-5482
REFERENCE 1 (bases 1 to 303750)
AUTHORS   Xu,J., Bjursell,M.K., Himrod,J., Deng,S., Carmichael,L.K., Chiang,H.C., Hooper,L.V. and Gordon,J.I.
TITLE     A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis
JOURNAL   Science 299 (5615), 2074-2076 (2003)
PUBMED    12663928
REFERENCE 2 (bases 1 to 303750)
AUTHORS   Xu,J., Bjursell,M.K., Himrod,J., Deng,S., Carmichael,L.K., Chiang,H.C., Hooper,L.V. and Gordon,J.I.
TITLE     Direct Submission
JOURNAL   Submitted (31-OCT-2002) Department of Molecular Biology and Pharmacology, Washington University in St. Louis, 660 S. Euclid, St. Louis, MO 63110, USA

FEATURES             source
     1..303750
     /organism="Bacteroides thetaiotaomicron VPI-5482"
     /mol_type="genomic DNA"
     /strain="VPI-5482"
     /db_xref="taxon:226186"
     complement(87..1649)
     /locus_tag="BT1210"
     complement(87..1649)
     /locus_tag="BT1210"
     /codon_start=1
     /transl_table=11
     /product="cytochrome D ubiquinol oxidase subunit I"
     /protein_id="AA076317.1"
     /db_xref="GI:29338517"
     /translation="MIESIDSLIDWSRAQPAWTHWIFVPLTGLAVVGMETUL
     YKGNFWEKTAQFVKMLFGINFAIGVATGLILEFFGNTNWSYFVGDFGAPLA
     IEGIAFMEATFIAVDFGWDKSKRFLASTWLTGLGATISAWWILVANWQHVP
     GMENPDTRVEMDFWAVKPFHTVLSGWLGVFVVGISCNFLKKRR
     EFALASIKIGALFGLVSLAVMTGDSGYOIACTOPMKLAAVGLVEGGTNGVLGI
     GMLNPKTYNDKDPLEFRIEIPMSLSELAERVDGVVPGITNIIEGGYQKDGTTA
     LSAEKIERGTAKALAAAYAAKASAGHEDAVAYQVLQENMEYFGYTKOVNQLV
     PNPLNFYGRFIVILGYSYFLFFIVLFFYIKDKLSKRMWMMHIALTLTPLAYIAGQ
     AGWVAECGCGFWAIRDMLPTSAAISKLDVGSVQATFFIILFTVMLIAEIGIMVKA
     IKKGPEEISNR"
     complement(1670..1906)
     /locus_tag="BT1211"
     complement(1670..1906)
     /locus_tag="BT1211"
     /codon_start=1
     /transl_table=11
     /product="hypothetical protein"
     /protein_id="AA076318.1"
     /db_xref="GI:29338518"

```


Db	247226	GTC-----	247224	gene
QY	646	ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGly	665	CDS
Db	247223	-----TACACCTACAACTGTGACGAGCATATTCG-----TTAAGT	247188	
QY	666	PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal	685	
Db	247187	TTTTCCTCCGCTCGGTATGCGGAATACGTTGCATGTCGCAAGAAACACTATATGCTT	247128	
QY	686	PheAla-----GlnAsnLysLeuAlaArgTyrGluThrArgThr	698	
Db	247127	TATGCGGAGTGGCAACTATTGCCCCGTCAGAACCGGTAGACCGTATGAGGACCGTACA	247068	
QY	699	ProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu	718	
Db	247067	CCGGGGGCTAACCTGTTTCATTTGGGTGGCTCGCTG-----AATATCCGGTTGGCAGG	247014	
QY	719	TrpAsnTrpTyr-----ValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAla	735	
Db	247013	ACTAACGAGATAGAAATAACCTGACCGCCCGTAATATCTTCAATACCGATATTACAAT	246954	
QY	736	HisSerSerPheLeuSer-----AspThrProGlnMetGlyArgSerPheThrGlyGly	753	
Db	246953	CATCTTAGTTTACAGGAAGTAGAAATTCGGAACCGGACGAACTCCAATTATTA	246894	
QY	754	ValAsnValLysPhe 758		
Db	246893	ATCAAAATTCATTT 246879		
RESULT 28				
AE006167				
LOCUS	AE006167	11549 bp	DNA	linear
DEFINITION	Pasteurella multocida PM70 section 134 of 204 of the complete genome.			
ACCESSION	AE006167	AE004439		
VERSION	AE006167.1	GI:12721638		
KEYWORDS	Pasteurella multocida			
SOURCE	Pasteurella multocida			
ORGANISM	Pasteurella multocida			
REFERENCE	1	(bases 1 to 11549)		
AUTHORS	May, B. J., Zhang, Q., Li, L. L., Paustian, M. L., Whittam, T. S. and Kapur, V.			
TITLE	Complete genomic sequence of Pasteurella multocida, Pm70			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	98 (6), 3460-3465 (2001)		
MEDLINE	21145866			
PUBMED	11248100			
REFERENCE	2	(bases 1 to 11549)		
AUTHORS	Zhang, Q. and Kapur, V.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA			
FEATURES	Location/Qualifiers			
source	1. .11549			
	/organism="Pasteurella multocida"			
	/mol_type="genomic DNA"			
	/strain="PM70"			
	/db_xref="taxon:747"			
	77. .391			
	/gene="PM1279"			
	77. .391			
	/gene="PM1279"			
	/codon_start=1			
	/transl_table=11			
	/product="unknown"			
	/protein_id="AAK03363.1"			
	/db_xref="GI:12721639"			
	/translation="MFAMLVEINIKEGKEQBEFLVFARNHIGTROEPGNLRFVLRDP RIKTRFYAYEVTVNEQALEAHKQTAHYHRCVAELEPLMTGARSKKLFDWVMPENLTNE CG"			
gene				
CDS				
384. .1073				
/gene="dod_1"				
/note="synonym: PM1280"				
384. .1073				
/gene="dod_1"				
/codon_start=1				
/transl_table=11				
/product="Dod"				
/protein_id="AAK03364.1"				
/db_xref="GI:12721640"				
/translation="MDKQLIQOLTKPTQPSVGLASDLWKFAFDTLTTLTSHHRLRLHF DIGQGFSPFFTGALAVKQFSPMLKVLHMVNDPFFHAKACADAGADITIQEQBT DCIAGTISYLRHYPLDILGLTCLPTEIDLLTPYLAQVDLIQILTLDRPTGVKAEQT AVIKRITRISNMLGHRDQKLIISVDGSMNLALASQLFPLGIDWVSSALFSQADVDA TLSEKTHLCR"				
complement (1122. .1997)				
/gene="PM1281"				
complement (1122. .1997)				
/gene="PM1281"				
/codon_start=1				
/transl_table=11				
/product="unknown"				
/protein_id="AAK03365.1"				
/db_xref="GI:12721641"				
/translation="MIKGILLSLASVLFGLYFSTLLQPLGGEDIFGFRIIFTPLF VIAVFLFKQAVGVYFVKIQKQPLLIIGFLFNAGMMGIIQLWFLWPNNGSALVS FYALLPLTMVLVGVKFIHVRKEHTSRVKFLAVTAAIGVFNILLKGLSWEALVSRGY STVFATRKILKINDLAGCLESLLPLVCIFYPAWQVDLQAVQANENILLLLVLLGLI SGVALNTYIVASSLLPINVLGUGTAEPIMLFSVFLICEQLDSEITPILFICLMISMI LFWSEGLVNLKRRYK"				
2144. .4480				
/gene="PM1282"				
2144. .4480				
/gene="PM1282"				
/codon_start=1				
/transl_table=11				
/product="unknown"				
/protein_id="AAK03366.1"				
/db_xref="GI:12721642"				
/translation="MLCKPNHGLLILISEVVFADQLDEIHVSVVNGKASLTALA SQVRISKLDNVKSGTIGETITNVSGVQNNYFGPNAGPMIRGLSGQVQIMQNDMP IHMAAVSSNLAIASNLASVKEIQIKSGVASVLYGKSGIGGAVNLTDIAIQAHTAEK AFRGVEWRAGLAPNTYQVQLQNNQNVFAYLIDASLSEISRRKIPGRSKDQICVDY EYLKWTKRLSQCVNIKLSINPKYFKYISQFYLDHYKNPDYGLSEGDOYTNMSAG GNPLNPEYGVSPYAEKPGLEEVYPVNGRIINAHSSTKQLTVGASVLDKGYTG ASFHNTRYGVGPGSYLASRAASQKSYLPVSVTNETNLNLKQLNALVMPISKVE IOLGQSRDKEWLASHLSNAFHTYNDYRSSFEQQLFNHHVLTGTVGEMARQKIFPS SGKDSYLPNNMDTKSLFILEKIDVSPFVVEFGQWQVKYVVDVTNNHSSRGKGGY SKNRTFFVONTHLGLLFPQVDTWYLLKQSRQWQERAPENVNELYADNNHYALLIEENG RLKKEQSDNWLSTGVNLEQFNAELTWFORNFDNFHYLGYTGISRAGLAVKEWQAPL RMNGVELDLYTQHTPHFGDLTHLFDYVHHKLTAPDHIOTKRIYGLPHLPTSRIG GDLTVQWRGRTSVTHVKOKRVTKVEVDDDELPAKKTFLVDLGLSYVYPLTRATLEI SLINQNLNREARLNNSPRLAPLPGRNVLGVKLMF"				
complement (4540. .5421)				
/gene="ilvy"				
/note="synonym: PM1283"				
complement (4540. .5421)				
/gene="ilvy"				
/codon_start=1				
/transl_table=11				
/product="ilvy"				
/protein_id="AAK03367.1"				
/db_xref="GI:12721643"				
/translation="MNFYTLKFLHLCETKNFAKTAONHMSPTLSRQIORMEODVG QILFLRNRQVELTDAGEKLLPEAQOSWMOQLKQQLRDSAQELNSELKLFCSVTA AYSHLPALLERFTRYPKVEIQLTTGDSLSALQVETQVDLAGQPKNLPHNIAFH YIDLSLIAPIACSATQLLQQTPIWQIQIPFLPVEGPARQIDQWFOQKTKHP KIYATVAGHEAIMPWALGCGVAMLPDVVIKNSPMNSQVSIHLKLTPTTTPFALGICVQ KKSLLQPLVRAFWQILQ"				
5674. .7149				
/gene="ilvc"				
/note="synonym: PM1284"				
5674. .7149				

		/gene="ilvc"		Pred. No.:		Length:		11549	
		/codon_start=1		Score:		354.50		Matches:	
		/transl_table=11		Percent Similarity:		38.78%		Conservative:	
		/product="ilvc"		Best Local Similarity:		21.97%		Mismatches:	
		/protein_id="AAK03368.1"		Query Match:		8.78%		Indels:	
		/db_xref="GI:12721644"		DB:		1		Gaps:	


```

/molecule="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="II"
/notes="biotype: El Tor"
/complement (90..1652)
/gene="VCA0052"
/complement (90..1652)
/gene="VCA0052"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95966.1"
/db_xref="GI:9657435"
/translation="MRLKPLSTSLSCVYRQATOPDGLVPEGEIIPNDOLPPDAREGL
QLNCKTLACNFCGLSEAKHVLORVNPKPAMVCRBCGAPFPPLANNRQVNLHLRL
QLHSDGLPACHNPACPTGLSVTHKQLYHAFVSGDRQRYRCKTCHSTFVDKWSGAN
HKLNFQBSLLGLFTGYSVREICRKLINPNTFYDHLDHIAASCRRLKAMIDARWNH
AKAYQAFASHYQLOAHSHNGVYWTADQGTGYILCOHLYNSADEQPTGSLDHPDYQT
PARFVPOQYAAEAQTHKALTLLRLERIDNRQYLILARANVEDPMGNLAQHYFPSKGA
VIRPPYTSYAHYHLVLMCDPCKRVSIFMPQDPLLRSAALSLSLRIRCONVLMVYE
BDPWOQSPLEKTDIVHMGWRDRWAIAHSGTKAGICVLAGSNPDPOHWOHASTIR
QIEYQNEFQLLESFINEPRKLRPGCIQLPLLDIFRAMEHLCYQDQKHGTAQNLGL
SQRLTKELLS"
/complement (1624..2361)
/gene="VCA0053"
/complement (1624..2361)
/gene="VCA0053"
/notes="similar to GB:U14003 SP:P09743 GB:M60917 PID:147309
PID:537224; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="purine nucleoside phosphorylase"
/protein_id="AAF95967.1"
/db_xref="GI:9657436"
/translation="MAKLNWRVTWATPHINAQPGDAETVLMPCDPLRAKYIETFLUE
DVKQCVDSVFGFTGYTKGKVSVMGHMGIPSCSYVHELIAEYGVKNIIRIGSGG
AVRDDVKLMDVVGNGASTDSKNVRIRPSGHDFADLADYDLLETAVNQARAQVPPKV
GNVFDADLYTPEPIFEKMKKIGILGVDMEAAGLYGVAADLGAARALTILTVDHILR
GEKJSSDRQKSFNDMMKVALETAINI"
/complement (2420..2599)
/gene="VCA0054"
/complement (2420..2599)
/gene="VCA0054"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95968.1"
/db_xref="GI:9657437"
/translation="MILYCTNIFVQNFPTPTTKRLAQRLPYQFQISSLSTAYPDKR
EYFVTLSSSLVDIG"
/2641..3609
/gene="VCA0055"
/2641..3609
/gene="VCA0055"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95969.1"
/db_xref="GI:9657438"
/translation="MLETMRMTPLKIGISACILGAKVRFQDGHKISHFVTNELDRIA
EFVSCPMGMGLPVPRTLRILISENIALVETKDSRSDHTAALESYSQOKVAELOH
AELGVYVCAKSPGEMRVKYKNGTEKDGVGVTILMERKMWLPFIEDGLNDP
VLRENYLTFVCLHDFVTSMGDAPSAKKIVZPHEYSYKLTLMHPNPQYALCRILVAVY
KOYLDSTFOETRLGMLAHRASRKNNTVNHMQGFYKRSLNKQKQAEIAQVIES
YVGBELPLAPITLLKHYLAUFPDADYLAQQRYLNPFOELKRYGL"
/3599..4396
/gene="VCA0056"
/3599..4396

```

```

/gene="VCA0056"
/notes="similar to SP:P33358 PID:405858 GB:U00096
PID:1789448; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MerR family"
/protein_id="AAF95970.1"
/db_xref="GI:9657439"
/translation="MVCDEKRYAIRVAEITGVKPVTLRAWORRYNLVOPDTEKGHR
LFTDQIEMIRQISWLAKVAGKVALQSGVSESESIPOPVQGLECEETLLTALA
ALORSKAEQIATVLKPEYPLSVQAOFPQVTEALERVKGLRSLQIGLFRITLMSKL
AFILDAENKAAVKGKLCISLDEAGSLNMLWALAWAENGHOVTLEAVDDIRGLEEN
PGLAQYQIALHAHRAHPAAQOSALASLQQQGEQCVLSNVLQQLQS"
/4406..5815
/gene="VCA0057"
/4406..5815
/gene="VCA0057"
/notes="similar to GB:K01299 SP:P00914 PID:147268 GB:U00096
PID:1651311; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="deoxyribodipyrimidine photolyase"
/protein_id="AAF95971.1"
/db_xref="GI:9657440"
/translation="MRLVWRERDLRSFNTALTAALNSGDPVAAVYIATPEOWHOHL
APLOADLWERLAELQOELALNVLPLYQOVADFOAAVAVSOLAKTINATOVLANRD
YELDEQQRDQLAQQLISQGIINWSAFDDKCVLPQGSVTRKQGEPEPKVTPFKRAMLTL
FOPVIGKRNPRVALNWPALAEVLVHPEQAFDYPIDSTPWAADFETVRAQDRFCR
FPPVDYHQARDPFAPEGTSLSPLAIGVLSARQCVALYHSESGEAGQVWLSE
LWRFYFHLVAIEPNLSKRSRDFVWGNLWNNERFQLMCEKGTGYIVDAAMRQ
LNOTGMHNLRMIVASFLTLDHIDMRGERYFMSRLIDGIDYAAANGGQWQCASTGC
DQQVFRIPNPVSQGEKPDNGDFIRRWPELRSVSSAVIHQPMTPYPAVNSVLYPARL
VHKQEREVTLRLYKTAKG"
/5824..6744
/gene="VCA0058"
/5824..6744
/gene="VCA0058"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95972.1"
/db_xref="GI:9657441"
/translation="MMWKVWITGALAVSWSINAAVFDLPBEGSSIVGNTQYHVKVEKG
ETLADIATKYDIDGLALMAANRGVDFPLPOEGFVSIIPAKIILPKVAYGIVNLAEGL
RLYFPEPNEGKVIHPVIGRIGRDTPTVMQTSISSKRYKPTWTPPASIRKEYKAGID
LPPVVPAGPENLGDYAMRLAYSGSEYLIHGKNGKPGVSGKVSAGISMDKQLEWLY
QQVERGEKVRILNEPIKVADEEEDRVFIEVHEPLRSNGVKNKLVVPEELSNWLQEHK
ISDAKARAALVLAQNGVPEIAPPQWETIAQ"
/complement (6830..7189)
/gene="VCA0059"
/complement (6830..7189)
/gene="VCA0059"
/notes="similar to GB:M13176 SP:P09461 GB:M28554 PID:150894
PID:150896; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="major outer membrane lipoprotein"
/protein_id="AAF95973.1"
/db_xref="GI:9657442"
/translation="MRASPARNCDDVNSVHILLFDKYRTSITPTVQVKFSPMNKMLIA
AAASSVLLIAGCAGSGDEATTAKMNEISTQVSELNSQVAALASKVDQAAEAKAAQEE
AARANERIDNTIAQSYTK"
/7386..7562
/gene="VCA0060"
/7386..7562
/gene="VCA0060"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95974.1"
/db_xref="GI:9657443"

```

/translation="WALSKGQPMQVDFSRQICHSIAAYLQHVVPYLPYGFWDY

7608..8966

/gene="VCA0061"

7608..8966

/gene="VCA0061"

/note="similar to GB:102123 SP:P25888 PID:560801 GB:U00096

PID:1651360; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="Arp-dependent RNA helicase, DEAD box family"

/protein_id="AAF95975.1"

/db_xref="GI:9657444"

/translation="WESTLQFKDGLDNRLKLNLAHYNFKQATEIQQAIPLTIAGRD

LIAASKTGKTLAFLVPLMLHSLKTKAFSAKDPRLGLILVPTRELAKVYGLSRMLG

GLSYATLITGGENEDOVKALARGPRIVATPGRLADHLDRSLFLGLETLVLDEA

DRMLDGFALKELRIRHNARHRRRTLFESATLHDADVNMAMELLNPEKRIAGVGS

ESBKDIQHFVLCDDHDKLELDRLDAEYRQVLIIFATRAUDTRUTEKLNQNLK

Alignment Scores:

Pred. No.: 1.73e-07 Length: 17725

Score: 290.00 Matches: 185

Percent Similarity: 37.60% Conservative: 129

Best Local Similarity: 22.16% Mismatches: 313

Query Match: 7.19% Indels: 208

DB: 1 Gaps: 41

US-09-936-377-2 (1-758) x AE004349 (1-17725)

QY	6	LeuYleProIleValLeuSerIleLeuLeuIleAenThrProLeuLeuAlaGlnAlaHis 25	
DB	11182	CIITCTCCAGTATACCTGCGGTACTAGCGTACTCGCGGAGGGTCTGCTCATGCC--- 11238	
QY	26	GluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys---SerArgPro 44	
DB	11239	GAAACTGAGCTAGCCACTACCAAGAAAGTGGTGTGACGGCCAAATCGCATTCAGAGCGCT 11298	
QY	45	ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspIleIleSerGlyAsp 64	
DB	11299	---TGTGAGAAAGTGGCGGCTCAGTGGCGAGTGTCTTGAAGGGGAA 11340	
QY	65	ThrLeu---ArgGlnIysAlaValAenLeuGlyAspAlaLeuAspGlyValProGlyIle 83	
DB	11341	ACCTTTGAAAGCAAGCAAGCAAGCGAGTTGTACGATGCATTGAATCAACAGCGCGGTGTC 11400	
QY	84	HisAlaSerGlnTyrGlyGlyAlaSerAlaPro-----ValIleArgGlyGln 100	
DB	11401	AGCGTGACT-----GGCGAGCGGAGCGCCACAAATATTACGATTCTGGGCATG 11451	
QY	101	ThrGlyArgArgIleLysValLeuAenHisHisGlyGluThrGlyAsp-----Met 117	
DB	11452	ACCGGCAACCGGATAGCGATTGTTCTGTGATGGCATTCAAAGTCCGCGATGTTATGCGCT 11511	
QY	118	AlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSer---GlnGlnVal 136	
DB	11512	GCAGATATTAAACCAAGTATGCGCGTAAACACATTTAGCGCTGAGCAATGTGAACACAGATC 11571	
QY	137	GluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAenValAlaGlyLeuVal 156	
DB	11572	CAAGTGGTGAGGCGCGGAGTTCACCCCTATACGTTCTGGTGGATGGTGGCGTCTG 11631	
QY	157	AspValAlaAspGlyIleProGlu----- 165	
DB	11632	---ATCATCGAAAGTAAAGCGCGGAGATTAITTTGTACCACCGGTATTATACGTGGAT 11688	
QY	166	---LysMetProGluAenGlyValSer-----GlyCluLeuGlyLeuArgLeu 180	
DB	11689	GCGCATTTGACTTACTCAGGATCAGCAACCGGTACCAAGGTAAACCGCGTTAGCCATG 11748	
QY	181	SerSerGlyAenLeuGluTyrLeuThrSer----- 190	
DB	11749	CGTCACGTTGATGGCGAGCGGCTGCTCACGATCGACTACTGGCAAGGTGAAGAGACACGT 11808	

QY	191	-----GlyGlyIleAenIleGlyLeuGlyLys 199	
DB	11809	AACCTTTAATCAAGATCTGTATACCGCGAAGTGGATGGTTTACAACTTGGGTTCTCCCAT 11868	
QY	200	AsnPheValLeuHisThrGluGlyLeuTyrArgLysSer-----GlyAsp 214	
DB	11869	CATTATTGGTTAAAC---GATGCGCTCGCTTTAAACACGCACCTAGAGTATTTTGATGAT 11925	
QY	215	TyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAla 234	
DB	11926	TATGCC-----AAGCGTCGCGAAGGC 11946	
QY	235	AsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSer 254	
DB	11947	ACCTCTTCTATCCAAAAAGATGACCAAGTGGATCTGGTACGCTTT-----TAGCAG 11997	
QY	255	AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla 274	
DB	11998	TATCAACGCGACCAACCGCTCTGGCCAGCGTGTGGTCCCGATTACAG-----GCT 12048	
QY	275	AspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu 294	
DB	12049	AATCTAAGTTGGATGGATACGCTGGAAGGAAATTTCTACTGGCGC----- 12093	
QY	295	LeuThrGluLysAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp 314	
DB	12094	---AGTACAGAAAAACATCACACAGAACCGTTTGTATGCGG----- 12132	
QY	315	AspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArg 334	
DB	12133	---AATGATCGGTCTGGCGTGGCATCTCTCTATCGT 12168	
QY	335	TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPhe-----GluAlaLeuArg 352	
DB	12169	CGTGAATCGCTGATGAA-----GGGTTCATGATGAGCGCTTAGGA 12210	
QY	353	ValHisLeuAenArgAsnAspTyrHisHisAspGlyLysAlaGlyAspAlaValGluAsn 372	
DB	12211	CCACGCTCAATGCG-----CAAAAGAGTGGCAACAGGTAGTGGCTGCATCAG 12261	
QY	373	PhePheAsnAenGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArg 392	
DB	12262	TTTGCTTACGGTATGAGCGTTGATGGTTCATGATTACAGCGTCCGAAATCGATTCGACGG 12321	
QY	393	LeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThr 412	
DB	12322	ATGGAATCTTCA-----GGTACGATTTACAGCGCGATGACCGCTTT 12363	
QY	413	SerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePhe 432	
DB	12364	GCTCTCGC-----CGTGAGTACCGTTTGGCGTCTAC 12396	
QY	433	GlyValGluGlnAlaAsnTyrAspAsnPheThrLeuGlyGlyValArgValGluLys 452	
DB	12397	GGACAGACAAATCTCTTATTAGCGATGGACGTGCGGTTTTCGTTTTGATGCG 12456	
QY	453	GlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAenTyrTyrLysGln 472	
DB	12457	CAAAATTCAGCCCTAAAAAT-----ACCGATCTATTATCGGCTACAAA--- 12501	
QY	473	ProLeuProAspLeuGlyValAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAen 492	
DB	12502	---GTGCTGACCATGGGGAGC---AGTGAAGTGGTCCACCGATGGCTCAATCTCTATCAA 12555	
QY	493	TyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuPro 512	
DB	12556	TGGCAT-----CCAGATGGAATTAATTACTGCTGACACACCGGCTTCGCGCGCA 12609	
QY	513	SerThrGlnGluLeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGlu 530	
DB	12610	AGTTATGACAAAGCGTATGGCGCAAGCGATCACAGTTTGTCCCGCTAACGCCATTATT 12669	
QY	531	Val---GlyAsnLysHisLeuAsnLysGluArgSerAenAsnIleGluLeuAlaLeuGly 549	

```

Db      12670  ATCAAGCCCAACAATAAGTTACGGCTGAACACAGCGAGTTCAGTTGGCCACAA 12729
Qy      550   TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
Db      12730  TAGCACAATGGTCAGACGCAATTTACGTGGCGGTGTTTATTCATCCATCTTCGATACTTT 12789
Qy      570   IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGlu 589
Db      12790  ATGTATGTCACAAAGTCGGTTATGACACAGCCACAGGTCGGTGATTCAGCAGTACCAA 12849
Qy      590   MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIle 609
Db      12850  -----AATATCGCGGGCGTCAAAACCTACGGGCTGAATGTCGTA 12891
Qy      610   TyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg 629
Db      12892  ATGCATCGTCTAGATGATCGTTGGAGTGTGAAATAAGCTCGGATACGTGGATGTA 12951
Qy      630   -----LeuLysAsnLeuProSerLeuProGly----- 638
Db      12952  GATGTGAAACCAATACGTGGCACCTTAACCCCTAGAGGGAGTGTGCAGCTCAAC 13011
Qy      639   -----ArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGln 654
Db      13012  TACCAAGAGAGCGATGGACGCTTAC----- 13038
Qy      655   AsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp 674
Db      13039  -----AGCCGCTGAACCTGG-----GCGAGTCCCATGAGC 13068
Qy      675   ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyr 694
Db      13069  CGAGTCCGACT-----TCACCAACAGACCAAGGCAAGAACTGAATG 13113
Qy      695   GluThrArgThrProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThr 714
Db      13114  GCCACG--ACCACAGTTGGTGAGTTGGACATCGGCTCAACTAC----- 13158
Qy      715   ArgTyrGlyGluTyrAsnTyrTyrValLysAlaAsp-----AsnLeuLeuAsn 730
Db      13159  -----CAATGAAGCCACACAGTGTGAGTCCAGTTTCAACGTGTGTTAACTGCTCGAT 13209
Qy      731   GlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro----- 744
Db      13210  CGAGATACACCGCTATCAAGATCGCTGCTGTGACCCCAAGTACACCCCTCTATTCC 13269
Qy      745   ---GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      13270  ACGGAGCCGGACGCTATTTCACC-----GTGCACGCTAAGTAT 13308

RESULT 30
AX079015 LOCUS AX079015 1173 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 26 from Patent WO0107587.
ACCESSION AX079015
VERSION AX079015.1 GI:13158590
KEYWORDS
SOURCE Pseudomonas sp.
ORGANISM Pseudomonas sp.
REFERENCE 1
AUTHORS Fraser,C.M., Venter,C., Tuemmler,B., Hoheisel,J., Duesterhoeft,A., Hilbert,H., Timmis,K.N., Moore,E., Stratsch,M. and Heim,S.
TITLE Dna sequences coding for iron and heavy metal transporters
JOURNAL Patent: WO 0107587-A 26 01-FEB-2001;
The Institute for Genomic Research (US) ; QIAGEN GmbH (DE) ;
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;
Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische
Hochschule Hannover (DE)
FEATURES
source
1..1173
/organism="Pseudomonas sp."

```

```

/mol_type="genomic DNA"
/db_xref="taxon:306"
BASE COUNT 236 a 378 c 362 g 196 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 7,28e-08 Length: 1173
Score: 271.50 Matches: 96
Percent Similarity: 42.15% Conservatives: 49
Best Local Similarity: 27.91% Mismatches: 148
Query Match: 6.73% Indels: 51
DB: 6 Gaps: 12
US-09-936-377-2 (1-758) x AX079015 (1-1173)
Qy 430 SerPhePheGlyValGluGlnAlaAsnTrp-----AspAsnPheThrLeuGluGly 446
Db 85 GCGCTCTTCATCTGAGGAGCAG-----TGGCAAGCCACGACGCTGAACTGAGCCTG 138
Qy 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArg 466
Db 139 GCGCGCGCATGGAG-----CACACCGCGCTAGACCCGAGCCAAAGCAAC 186
Qy 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSer 486
Db 187 GAGAAGCTTCGTCTGAC-----GCCGACAGCCGACGAGCTTCAACGCC 228
Qy 487 PheAlaLeuSerGlyAsn-----TrpTyrPheThrProGlnHisLysLeuSerLeuThr 504
Db 229 TTCAGCTGTCTGCGGGGGCTGTGTACACAGCTCGACCAATCTGTCGTGGCGGCCAAC 288
Qy 505 AlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisVal 524
Db 289 CTCGGCTACACGAGCGCGCCCTTACCTTTTACGAGCTGTACGCCAATGGCGCGCACGTA 348
Qy 525 AlaThrAsnThrPheGlyValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIle 544
Db 349 GCCACCGCGCGCTTTGAAGTCGTGACGCCACGCTGAACAGAAAGGCCATTTCCGCC 408
Qy 545 GluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsn 564
Db 409 GACCTGGCGCTGCGCTTTCAGAACCGCCACAGCGGCGAGTGTGGTGTGTTCTACAGC 468
Qy 565 ArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer 583
Db 469 CACTTTTCGCAACTACATCGGCTGATTGCGCACCAACACCGCGAAGCGCAGCATGAC 528
Qy 584 IleGluAspAspSerGlu-----MetLysLeuValArgTyr 595
Db 529 CATGAGGAGGAGGATCAGCATCATGATCATGATCATGAGGGCTTCCCGGAATACCAGTAC 588
Qy 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
Db 589 CAAGCGTCGCGCGCGCTTCTACGCGCATCGAGGCCAGGACCGCTGCGGTGGCCGAG 648
Qy 616 Arg-----TyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLys 631
Db 649 AACCGTTACGCGCAGCTTCGCGCTGGAGTTGTCGGTGACTATACCGAGCC-----AAG 702
Qy 632 AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGln 651
Db 703 AACCTCGACAGT----- 714
Qy 652 AspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSer 671
Db 715 ---GCGCAGCACCTGCCCCCGCATCGCCCGCTGCGCTGAACAGCGGCTGTCTCTGGAG 771
Qy 672 LeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeu 691
Db 772 CTG---GACCGCTGCGAGCGCGGTARACGTCGAGCAGCGCTGCGTCGACGACCGCAAA 828
Qy 692 AlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArg 711

```

```

Db      829  CCGCCACGACGACCGAGTGGTACACACCTTGGCGCGGCGGTGGCTACCGC 888
Qy      712  ArgAsnThrArgTyrGlyGluThrAsnTrrTyrVallysAlaAspAsnLeuLeuAnGln 731
Db      889  TTCGACATTCGCGAGCAGTGGCTGGCGTTTGTGCGCGGAGAAACCTGACCGACGAG 948
Qy      732  SerValTyrAlaHisSerPheLeuSerAsp---ThrProGlnMetGlyArgSerPhe 750
Db      949  ACCGTGGCTATCGCAGTTCGCTTGGCGATACGACCGCGCGCGGCGGTACGGT 1008
Qy      751  ThrGlyGlyVal 754
Db      1009  GAAGTCGGGTG 1020

RESULT 31
LOCUS   AE005425                10161 bp    DNA        linear    BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 44
of 290.
ACCESSION   AE005425
VERSION     AE005425.1
KEYWORDS    GI:12516166
SOURCE      Escherichia coli O157:H7 EDL933
ORGANISM    Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 10161)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
MEDLINE
PUBMED    11206551
REFERENCE  2 (bases 1 to 10161)
AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
LOCATION/Qualifiers
1. 10161
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
172..247
/gene="asnV"
/note="synonym: Z3149"
172..247
/gene="asnV"
/product="tRNA-Asn"
/note="anticodon: GTT"
/function="RNA; tRNA"
complement(303..1235)
/gene="erfK"
/note="synonym: Z3150"
complement(303..1235)
/gene="erfK"
/note="Residues 1 to 310 of 310 are 99.03 pct identical to
residues 1 to 310 from Escherichia coli K-12 Strain
MG1655: B1990"
/codon_start=1

FEATURES
source
1. 10161
/translation="MRRVNIICSFALLFASHTSLAVTYPLPPGSRSLVQGLTVPD
HTNPLETFAAYCGQGLSNMLEANPGADVFLPKSGSOLTIPOQLILPVTVRKGLVWV
AEMRLYYPPDSNTVEFPFIIQAGARETFRNWTTVRKEQEAFTPTPTNRREYAK
RGESLPAPFPDPNPMGLYAIYIGRLYAIHGTNANFGILRVSGGIGIRLNDIKYL
FDNPVGTFRVQIIDQPVKTYTTPDGSKWLEVHEPLSRNRAEVESDRKVPLPVPFSLRA
FINQGEVDVNRANAALQRSRMPVQISSGRQMF"
/gene="cobT"
/note="synonym: Z3151"
complement(1300..2379)
/gene="cobT"
complement(1300..2379)
/function="enzyme; Biosynthesis of cofactors, carriers:
Cobalamin"
/note="Residues 1 to 359 of 359 are 95.82 pct identical to
residues 1 to 359 from Escherichia coli K-12 Strain
MG1655: B1991"
/codon_start=1
/translation="MOTLANLLNTIPADPAAMRAQRHIDGLLKPVSGSLGRLEALGV
QAAGMPLNGIPGHVKAIVMCDHGVMEEXVAISPKEVTAIOAENNRGTGTGVCVL
OLAAGANVHIDVDITDAPIPGLINMRVAXSGNISAPAMSRQAEKLLDVTICVT
RELAKNGVTLFGVGLGMANTTPAAATVSTITGRDPERVVGIGLANLPDCLKANKIDVV
BRAITLQNPQDQGVNVLAKVGGFDLVGMAGMLCAASCGPLVLLDGLFSLAAALAAC
QMSPAIKFPLPSHLSAGKARIALSHGLEFFLLNMDRLGSGAALAMPFIEAAC
LYNNNGELAAASKIVLPNGNTSDLNS"
/gene="cobS"
/note="synonym: Z3152"
complement(2391..3134)
/gene="cobS"
complement(2391..3134)
/function="enzyme; Biosynthesis of cofactors, carriers:
Cobalamin"
/note="Residues 1 to 247 of 247 are 99.59 pct identical to
residues 1 to 247 of 247 from Escherichia coli K-12 Strain
MG1655: B1992"
/codon_start=1
/translation="MSKLPWAMLSPTIRLPVPRWSQGLDFEHSRGIITPPLIGLL
GAISGLVFWLQAWGAPLALFVSLVLTMTGPHLDGLADTCGVFSARSRRMLE
IMRDSRLGTHGLALIFVLAKILVLSLALRGEFPLASLAAACAVSGTALLMYRH
RYAREGLGNVFIGIDRGTCVTGLGLAAIFAAVLPGMHGVAMVVTVAIFILGQL
LKRTIGGOTGDTLGAAILBELGVLFLALL"
/gene="cobU"
/note="synonym: Z3153"
complement(3131..3676)
/gene="cobU"
complement(3131..3676)
/function="enzyme; Biosynthesis of cofactors, carriers:
Cobalamin"
/note="Residues 1 to 181 of 181 are 100.00 pct identical
to residues 1 to 181 from Escherichia coli K-12
strain MG1655: B1993"
/codon_start=1
/translation="WMILVTGABSGKSRHAEALIGDSQVLYIATSOILDEMAARI
EHRQGRPEHRTVERWQHLDELINPNEVLECCVTMTNLLFDYGGDKDPDE
WDIQAEQAINAEIOSLIAACRCPAKVVLVTNEVGMGIVPESRLARHFDYIAGKVN
QLAAAAAEVNLVVSIGVKIK"

```

```

misc_feature
3945..6395
/note="O-island #81; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
4017..4418
/gene="Z3154"
4017..4418
/gene="Z3154"
CDS
4017..4418
/function="orf; Unknown function (Insertion Sequence
Associated)"
/note="Residues 1 to 133 of 133 are 100.00 pct identical
to residues 1 to 133 of 133 from GenPept 118 :
gi|3414881|gb|AAC31492.1| (AF071034) L0013 [Escherichia
coli]"
/transl_table=1
/product="unknown protein encoded by ISEc8"
/protein_id="AAG57055.1"
/db_xref="GI:12516171"
/translation="MGTKVSDMQKNVTGRRKGCNPVPFQKQVLAASCEPGISISK
LALENGINANILFKWQRMREGKLLLFSSPQLLPVTLDAABQPSLEADPETUSI
SCVTFRHGTRFNGNVSEKLLTLIIQLKR"
4415..4762
/gene="Z3155"
4415..4762
/gene="Z3155"
CDS
4415..4762
/function="orf; Unknown function (Insertion Sequence
Associated)"
/note="Residues 1 to 115 of 115 are 100.00 pct identical
to residues 1 to 115 of 115 from GenPept 118 :
gi|3661485|gb|AAC61716.1| (AF081284) Hp4 [Escherichia
coli]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by ISEc8"
/protein_id="AAG57056.1"
/db_xref="GI:12516172"
/translation="MIPDPSGKTLWLVAGITDMRNGPGLAAKVQTKLKDPMGSHVF
IPRSGSQVLLSTGDLCLLTKLRLGRFANPSARDGKVLFTQALMLLSDW
RQPKRLLSLTNL"
4812..6350
/gene="Z3156"
4812..6350
/gene="Z3156"
CDS
4812..6350
/function="orf; Unknown function (Insertion Sequence
Associated)"
/note="Residues 1 to 512 of 512 are 100.00 pct identical
to residues 1 to 512 of 512 from GenPept 118 :
gi|3414883|gb|AAC31494.1| (AF071034) L0015 [Escherichia
coli]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by ISEc8"
/protein_id="AAG57057.1"
/db_xref="GI:12516173"
/translation="MNDISSDDIFLLKORLAPQELIHAIQELKSNRREIDHQAOL
DKLRNFGSRKYSRIAQMEADNLRLQKESDITGRVYDPVAVRFLQTRTKPF
PSLPRDEKRLPAPPCPCNGCSLYLGEDTAELMRSAFVIRTVREKHACTQC
DAIQAAPSRPIERGLAGPLDLKVLTSKYAETHPLVROSEIYGOGVLLRSLISG
WDAACRLLSLEALHGVMTDGLHADDTPQVLLPQNKTKTGLWVVRDENA
GSALAPVAFVSPDRKHINPQTHLACFSGLVQADAYAGFNLVNGGITAEACWAHA
RKTHDVHVRISALTEALEQIGLYALEADIRGMPAPORLAERQRTKPKLSELES
WLREKNTLSRSESLAKAFAYALNQWPAITYANDGWVEIDNNIAENALRAVSLGRKN
FLFFGSDHGCGGALLYSLTGCKLNDVDPESYLRVLGVADNPVNRVSELLPWRIA
LPAE"
7787..9946
/gene="Z3159"
7787..9946
/gene="Z3159"
CDS
7787..9946
/function="putative membrane; Transport of small
molecules; Cations"

```

Alignment Scores:
Fred. No.:

2.79e-06 Length: 10161

```

Score: 265.00 Matches: 176
Percent Similarity: 38.16% Conservative: 122
Best Local Similarity: 22.54% Mismatches: 338
Query Match: 6.57% Indels: 145
DB: 1 Gaps: 40

US-09-936-377-2 (1-758) x AB005425 (1-10161)

QY 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
Db 7841 CTGTATTATGATTTCTGGGATAGTTTCTACTGCAACCGCAACTTCTTCAGAAACAAA 7900
QY 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db 7901 ATCAGCAACGAGAGAGCGCTCGTGCACCAAGATCGT-----TCGCAAGCAAC 7951
QY 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
Db 7952 CTTTGGAAAGCCCGCGACTATACAGGTATTGACCAACAACATTCGAGAACTCCACC 8011
QY 71 ---ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGly 89
Db 8012 AATGCTCCATAGCCGATATTTGCAGGACATCCCGGAGTAGAGATAACAGCAACTCC 8071
QY 90 ---GlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeu 108
Db 8072 TTGGCAGGCGGTAAACAAATCCGATTCGGTGGGGAAGCATCTCCCGTGTATTATCTC 8131
QY 109 AsnHisHisGlyGlu---ThrGlyAspMetAlaAspPheSerProAspHisAlaIleMet 127
Db 8132 ATTGATGTCAGGAGGTAATCTTATCAGCGCGCGGAGATAAATTATGCTGTGGGACTGTTG 8191
QY 128 ValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyr 147
Db 8192 ATAGATGAGTCTCGCTGGAGCGTGTGAGGTAGTGAAGGTCCATATTCCGCTACTGTAC 8251
QY 148 SerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMet 167
Db 8252 GGTTCACAGCAATTGGCGGTATTGTTAACTTCATCACCAAAAAGGAGGTGACAACTT 8311
QY 168 ProGluAsnGlyValSerGlyGluLeuGlyLeuArgLysSerSerGlyAsnLeuGluLys 187
Db 8312 ---GCATCTGGAGTTGTGAAGCTGTTTATAATTCCGCAACAGAGCTGGGAGAAATCA 8368
QY 188 LeuThrSerGlyIleAlaSerIleGlyLeuGlyLysAsnPheValLeuHisThrGluGly 207
Db 8369 ATCCCGCTCAGGGG---AGCATCGGT-----GGATTGATTATCCCATCAACGGT 8416
QY 208 LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db 8417 AGTTATTCTGATCAGGCAATCGTGATACGCCGAT-----GGACGCTCTGCCG 8464
QY 228 AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db 8465 AATACC-----AATCATGTAACAATAGTCAGGGTGTATGGTGGGTATAAATCCCGA 8518
QY 248 PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db 8519 AACCATCGT-----TTTGGCCTCTCGTTGATCGCTACAGACTCGCGACACAACTTAC 8572
QY 268 GluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr 287
Db 8573 ---TATGAGGATCCAGCGGAAGTTAT-----GAGGCATTT 8605
QY 288 LeuGlnLeuTyrProHisIleLeuThrGluGlu-----AspValAspTyr 302
Db 8606 AGTGTCAAAATACCTAAACTTAACAGAGAGAAAGTTGGGGTATTCTATGACACAGAGCTG 8665
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
Db 8666 GACGGTGACTATCTAAAA-----AAAATTTCATTTCCGACGCGTAT 8704
QY 323 AsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342

```



```

JOURNAL      DNA Res. 8 (1), 11-22 (2001)
MEDLINE      21156231
REFERENCE    11258796
AUTHORS      5 (bases 1 to 286857)
              Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
              Hayashi, T.
TITLE        Direct Submission
JOURNAL      Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
              Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
              Japan [E-mail: ken@gen-info.osaka-u.ac.jp,
              URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
              Fax: 81-6-6879-2047]
COMMENT      genome project.
FEATURES     Location/Qualifiers
             1..286857
             /organism="Escherichia coli O157:H7"
             /mol_type="genomic DNA"
             /strain="O157:H7"
             /sub_strain="RIMD 0509952"
             /db_xref="taxon:83334"
             63..1508
             /gene="ECs2545"
             63..1508
             /gene="ECs2545"
             /note="similar to YEBU ECOLI gi|1788140 percent identity
             96 in 481 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /product="putative rRNA methylase"
             /protein_id="BAB35968.1"
             /db_xref="GI:13362013"
             /translation="MLVQHTVYFPDAFLTQNRMPSTLSDFLAACQPLRRSIR
             VNTLTSVADFLRLTAPYQWTLPWPCEEGFWIEDDEDALPLGSTAEHLGSLFYIQ
             EASMLPVAAALFDAGNAPQVMDVAAAPGSKTQTQAAEMNEGAILANEFASRVKVL
             HANISRGSTFVDFGRVFGAALPFMDAILLDAPCGEGVVRKDPDALKNWSP
             SNOETAAORELIDSAFHRLRSGGTVSTCTLNRENEVCLMLKETYPDVAAGVFLP
             QFLPCKANKLFEGLHVFPQIYDEGFFVARLKTQAI PALTPKVKVGNFPPSPV
             KREARQAPQAAAGLWDGNLRLWQDKVLPVGVLEALIGKRESRLGKLAET
             HNKGVQHEAVIALSPDENAFELTQBAEEWYRGDRVYQQAAPVADDVIVTFQHQ
             PGLKRWIGSRUNSYRELVRDGLKFTSNA"
             1611..1862
             /gene="ECs2546"
             1611..1862
             /gene="ECs2546"
             /note="similar to B1836 ECOLI gi|1788141 percent identity
             100 in 83 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35969.1"
             /db_xref="GI:13362014"
             /translation="WESTGKTSVRIGAFEDIDDELHGHESPGDRLTLPCKSDPDLCLM
             QLDAMDATSIPALLNGHESVLYTRYDQDSAWIMRLA"
             1883..2158
             /gene="ECs2547"
             1883..2158
             /gene="ECs2547"
             /note="similar to B1837 ECOLI gi|1788142 percent identity
             98 in 91 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35970.1"
             /db_xref="GI:13362015"
             /translation="MAGILSWLFFPRCKISPKLNGTAPHPGDEMFAVLVFCVLDGGCE
             DIVDVYNTFQQCLYSMDQIRHGGCFPIEDFIDGFWRPAQEGDF"
             complement (2159..2818)
             /gene="ECs2548"
             complement (2159..2818)
             /gene="ECs2548"
             /note="similar to PP8A ECOLI gi|1788143 percent identity
             98 in 219 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="phosphoprotein phosphatase 1"
             /protein_id="BAB35971.1"
             /db_xref="GI:13362016"
             /translation="MMKQAPVYQRIAGHQRHILWLSGDIHGLCQLRKLWHCRFPD
             WDLITSGDVDRGQSLQCLQLLEQHWVRAVRGNHEQMAMDALASQMSLWLNKGG
             WFLADLNQOAKTALCKCOHLPFELVHSRTGKHTAHADYDDVYEWOKVDLH
             QVLWSRSLSERQKGGITGADHFWGHTPLHRVDIGNLHYIDTCAVFGGELTFVL
             Q"
             complement (3211..3552)
             /gene="ECs2549"
             complement (3211..3552)
             /gene="ECs2549"
             /note="similar to B1839 ECOLI gi|1788144 percent identity
             98 in 113 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35972.1"
             /db_xref="GI:13362017"
             /translation="MMKKSILAFLLITSSAALAAPOVITVSRFEYKDKWAENREEV
             MLTRCPGNALVINPSTLYQVPLNDVAQKVASGKTQAQPISVIQIDDPNPGKMSL
             APFIERAEKLC"
             complement (3565..4437)
             /gene="ECs2550"
             complement (3565..4437)
             /gene="ECs2550"
             /note="similar to B1840 ECOLI gi|1788145 percent identity
             98 in 290 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="putative resistance protein"
             /protein_id="BAB35973.1"
             /db_xref="GI:13362018"
             /translation="MLAFTWIALRPIHFTSLMLVFGFMYGAWLAPLITRLTLKRL
             RLQOHAAVNSLISATAMLVQGGIMGTGKGVFSPNMQVLTQFGVWLWQIVLAL
             VTLIVALQPNRPRLIFMLTQAGFILLAGVHATLNEGVTAKIHQTNHAILICAA
             WFGGLLPVLMCMOLIKGRWHQAIQALMRFSWCGHFAGVIGLASVNLALLITGFPPT
             LTYWGLLILKAILVMIMVIALANRYVLVPRMQDEDEAAFPWFVMTKLEWAGAV
             VLIISLLATLEFP"
             complement (4441..4815)
             /gene="ECs2551"
             complement (4441..4815)
             /gene="ECs2551"
             /note="similar to B1841 ECOLI gi|1788146 percent identity
             99 in 124 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35974.1"
             /db_xref="GI:13362019"
             /translation="MASTASLRYALAILTSLVTPSVWAHAILTHQYPRANAQVTA
             PQAILNPFSGVETGFGCAKITGPKNENIKTLPAKNEQDKRLIPLADSLKGYTYT
             VDMHVSVDGKTKGHYTFSVK"
             4954..5184
             /gene="ECs2552"
             4954..5184
             /gene="ECs2552"
             /note="similar to HOLE ECOLI gi|1788147 percent identity
             100 in 76 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="DNA polymerase III theta subunit"
             /protein_id="BAB35975.1"
             /db_xref="GI:13362020"

```



```

Db 213177 AACATCGT-----TTGGCTCGCTGATCGTACAGACTCGCGACCAACTTAC 213230
QY 268 GluTyrAspAspCysHisAlaAspIleTyrPheGlnLysSerLeuIleAsnLysArgTyr 287
Db 213231 ---TATGAGGATCCAGACGGAAGCTAT-----GAGGCATTT 213263
QY 288 LeuGlnLeuTyrProHisLeuLeuThrGluGlu-----AspValAspTyr 302
Db 213264 AGTGTCAAAATACCTAACTTGAACGAGAGAAAGTTGGGGTATTCTATGACACAGACGTG 213323
QY 303 AspAsnProGlyLysSerCysGlyPheHisAspAspAlaHisAlaHis 322
Db 213324 GACGGTACTATCAAAA-----AAAATTCAITTCGACGCGTAT 213362
QY 323 AsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
Db 213363 GACGAGACCATCCAGCGCAATTTGCCAAC-----GAAGTAAACACGACAGCCT 213413
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHis 362
Db 213414 GTTCCAGTCCGATGATTCAGCTCTGACCGTTTCAAT---AACAGACTGAC---ACCCAT 213467
QY 363 AspGluLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGlnAsn 380
Db 213468 GATAAGCAATACACTCAGCGGTCAATGTCAGAGTCACTTTTCGCTCGCTGCTGCTAATAAT 213527
QY 381 AlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGln 400
Db 213528 GAACCTGTGTACC-----GGTGCACAG 213548
QY 401 Tyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVal 416
Db 213549 TACAAAACAGACAGGTCAGCAAGAGTCCGGTGCATGACCTCAAGCAAACTCTGTACC 213608
QY 417 LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGln 436
Db 213609 GGC-----TTCAATTAATGAAGAAACAGCACTCGCTCTCTATTATGATGACGAGCA 213659
QY 437 Ala-----AsnTrp-----AspAsnPheThrLeuGlu 445
Db 213660 AGTACAGTCTCACTATTCCACAAATGACTCGCGATTCCGCGATCACTGACGATGGACA 213719
QY 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAsp 465
Db 213720 ATGGAGATTCGC-----CAATACTGGCTTCTTCAAGTGTGACCGCGTGT-----GAC 213767
QY 466 ArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla--- 484
Db 213768 GGAGTATCATATACCGCAGCATTAATAGCGATATCACTCTCTTGGCAGAGAGTCTCGAGT 213827
QY 485 ---ArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeu 503
Db 213828 GATCAGAAATGGTAACATCTCAACGCTCGCCTATTGAGTTCGATAACTTGGAGTTA 213887
QY 504 ThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHisGly 521
Db 213888 CGCGCTGCGTTCGCGCAGGCTACGTATTTCACACACTCTCCGACTTTTATGACAGACA 213947
QY 522 LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
Db 213948 TCTCGCGCGCGCATGTCCATAC-----GAAATCTCTGTATCTTAAGCTGAAACACTCC 214001
QY 542 AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeu 561
Db 214002 AATACTTTGAATTAGTGCAGATATATGTTATGATCATGCTGCTGATGACAGCGCAGTT 214061
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db 214062 TACTACTCAGAGCTAAAGATATATTT---GCAAGTCTGATCTGTGATGGCAGTAGTATT 214118
QY 582 LysSerIleGluAspSerGluMetLysLeuValArgTyr-----AsnGlnSerGly 599

```

```

Db 214119 TGCAATGGTAACACCAACTCTCCCGTAGTACTATTATTATGACAATATTGATCGG 214178
QY 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle 619
Db 214179 GCAAAACATGGGACTGGAA-----ATAAGCGCGAATATATATGCG 214220
QY 620 GlyValSerGlyAspTyrValArgGlyArgLeu----- 630
Db 214221 TGGGTTTTCTCGCCATATATCATGTCGCAATTTAATTCGTCGCAATATGAAACTTCAACA 214280
QY 631 -----LysAsnLeuProSerLeuProGlyArg----- 639
Db 214281 TTAATAAACCACTAATAACAGGAGAACAGGATAAACAGGACGTATAGGGCTGAAACATACT 214340
QY 640 -----GluAspAlaTyrGlyAsnArgProPheIleAla 650
Db 214341 CTTGTGATGGTTCAGGCCAACATACTCTGATGTTTTTATTTCGTCTGCTCGCTTAGTGCA 214400
QY 651 GlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPheHis 667
Db 214401 AAAGATGACAGTAAACGTCACCGAAACAAATGTTCCGGGCTGGGCCACTCTCAACTTTTGA 214460
QY 668 LeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal---Phe 686
Db 214461 GTAAATACAGAAATC-----GGTAACAGAGATCAGTACCGGATTAACCTTA 214505
QY 687 AlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly-Hi 701
Db 214506 GCATCAATCACTGACAGACAAACGCTACCGTACAGACATGAAACTATTCTCGCAGCA 214565
QY 701 sHisMetLeuAsnLeu-----GlyAlaAsnTyrArgAsnThrArgTyrGly 717
Db 214566 GGTTTTAATCAGCTATAGGTTTTTGTATGGAATTTCTGTGACAAAACAAATCGTTATGG 214625
QY 717 Y 717
Db 214626 A 214626

```

```

RESULT 34
AF337562 2589 bp DNA linear BCT 29-MAY-2001
LOCUS
DEFINITION Escherichia coli clone RDAL:ex unknown gene.
ACCESSION AF337562
VERSION AF337562.1 GI:14211709
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 2589)
Allen,N.L., Hilton,A.C., Betts,R. and Penn,C.W.
Use of representational difference analysis to identify Escherichia
coli O157-specific DNA sequences
FEMS Microbiol. Lett. 197 (2), 195-201 (2001)
JOURNAL MEDLINE
PUBMED 21213883
REFERENCE 11313134
2 (bases 1 to 2589)
Allen,N.L. and Penn,C.W.
Direct Submission
AUTHORS
TITLE Submitted (18-JAN-2001) School of Biosciences, University of
Birmingham, Birmingham B15 2TT, UK
JOURNAL
FEATURES
Location/Qualifiers
1..2589
/organism="Escherichia coli"
/mol_type="genomic DNA"
/serotype="O157"
/db_xref="taxon:562"
/clone="RDAL:ex"
471..>2589
/note="similar to TonB dependent outer membrane proteins"
/codon_start=1
/transl_table=11
/product="unknown"

```

```

CDS

```


TITLE	Direct Submission
JOURNAL	Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada
COMMENT	<p>-----</p> <p>This represents the February 3, 2003 version of the continually updated, reviewed, <i>Pseudomonas aeruginosa</i> PAO1 genome annotation, from PseudoCAP (see http://www.pseudomonas.com for latest updates, and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.</p> <p>'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.</p> <p>Class 1: Function experimentally demonstrated in <i>P. aeruginosa</i>.</p> <p>Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).</p> <p>Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.</p> <p>Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.</p> <p>-----</p> <p>Location/Qualifiers</p> <p>1..16941</p> <p>/organism="Pseudomonas aeruginosa PAO1"</p> <p>/mol_type="genomic DNA"</p> <p>/strain="PAO1"</p> <p>/db_xref="taxon:208964"</p> <p>74..1009</p> <p>/locus_tag="PA3264"</p>
FEATURES	
source	
gene	

```
CDS
74..1009
/locus_tag="PA3264"
/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG06652.1"
/db_xref="GI:9949389"
/translations="MRALTALSRFVSASFFVWLLFAIAFFRPAHFLPLTAIAPIALLGVFMGMLTKGEDFEFRVAHRPVRVLIGLAQFVINEGLAWLRCRLQLPAEIAIAGVILMPCPGTASNVTMLRGDVALSVYAITSTLPALVPALWLLASAMLVAFVIAIMQSITQVVLYPALGLAQTLBERTROVAEVLPLVSVFSIVTWIIAAVVASQARIABESGLIAMVLMHGFGILLGYLTGKLTGMPLAQOKALAEVMQMNSGLGAALANAHESPLAAVPSALFSWHNLSGSLAALFRRLDDTPR"
1149..1463
/locus_tag="PA3265"
1149..1463
/locus_tag="PA3265"
/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG06653.1"
/db_xref="GI:9949390"
/translations="MSWILLFFAGLFGEVGNWAVKLTGEGSKPLPTVTLTAMLNVISGLLGLAMKHPLPTGAYAVMTGVGAGTVIGIVLFGESMALLRLASVALIICVGLGLKLSH"
1678..1887
/gene="capB"
/locus_tag="PA3266"
/note="synonym: cspa"
1678..1887
/gene="capB"

gene

CDS
1149..1463
/locus_tag="PA3265"
/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG06653.1"
/db_xref="GI:9949390"
/translations="MSWILLFFAGLFGEVGNWAVKLTGEGSKPLPTVTLTAMLNVISGLLGLAMKHPLPTGAYAVMTGVGAGTVIGIVLFGESMALLRLASVALIICVGLGLKL
```



```

Db      5559  GGGACCGCGCAACGGCTGGC-----GTGGCGCTGCTGATTTCGGGGTG----- 5512
QY      211  LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro 230
Db      5511  AAGGGCGCGCACTAC-----
QY      231  ArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArg 250
Db      5496  ---CGGACGGCAACACGACACGATATCGACGAGTGTCTCGAACCCATTGGCAG 5440
QY      251  ArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp 270
Db      5439  CTCACCGACGCGCAGCAG-----TTGGGGCGCAACTTCCTCACTACTAC--- 5398
QY      271  AspCysHisAlaAspIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeu 290
Db      5397  GAGCGCTACGCGCATATG-----
QY      291  TyrProHisLeuLeuThrGluAspValAspTyrAspAsnProGlyLeuSerCysGly 310
Db      5379  ---CGGGCGGCTGACCCAGCGCGAGTACGAC--GACGATCCCTTCAGTCCGTA--- 5329
QY      311  PheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyrIleAspLeu 330
Db      5328  ---CGGACTGGGACAATTTC 5311
QY      331  ArgAsnLysArg-----
Db      5310  CGTGGTCGGCGCAAGACTTCTCGTGAAGTACACCCGCGCAGGTGCGACGCTCACCCAG 5251
QY      335  TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
Db      5250  TTCCAGGTGCTCACTACTACAGCAGCAGTTCCTCCGCGCAGCAGCATCGCGCGGCAAC 5191
QY      355  LeuAsnArg-----AsnAspTyrHisHisAspGluLysAlaGlyAspAla 369
Db      5190  CTCAGGACCATCACTCGTACCCGCGGCTACCATGTGTTCCGGTGGAGCGCGGTGTC 5131
QY      370  ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro 389
Db      5130  TCGCGGATCTCTTCGCGCGCGCGCACCC-----CAGGAG 5095
QY      390  IleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db      5094  GTCCGC-----ATCGGCTACCGCTACCTCAAGGAA----- 5065
QY      410  SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn----- 424
Db      5064  ---GCGATGAACGAGCGCGCCAGCTGGCCCTGTGTCGACAAATGTGCGCGCGGTGCT 5008
QY      425  -----LysValGlnHisTyrSerPhe 431
Db      5007  CCCGGCTCGACCGGCATACCTACAGGACCGTACCGCGGGTACCGAGGCGCAGCGCTTC 4948
QY      432  PheGlyValGlnGlnAlaAsnTyrAspAsnPheThrLeuGluGlyGlyValArgValGlu 451
Db      4947  TACATCGACACAAAGATCGATGTCGCAACTGACCATCACCCCGCGCATCGCTCGAG 4888
QY      452  LysGlnLysAlaSerIleArgTyrAspLysAlaLeuLe----- 464
Db      4887  AAGATCGACAGCACTGGCGC---GACCGCGCGGTGCTCGGCTGACGCGCAAGCGGTG 4831
QY      465  -----AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481
Db      4830  CAGGAGAGAGCGCAGCAGGACTACACGAAACCGTTCGCGCGCGCTG----- 4783
QY      482  GlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeu 501
Db      4782  ---AGCCTGATGATCACTTTCAGCGACGAGTGG----- 4753
QY      502  SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGly 521

```

```

Db      4752  AAGCTGTTCGCCCACTACTCCGAATCGTTCCGCGAGCGTCGAC---TACTTCCAGCTCGGC 4696
QY      522  LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
Db      4695  CAGGGCGGACGGGCACTGATACCGCGCGCGC-----CTGAGCGCGGAGAGGCC 4645
QY      542  AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeu 561
Db      4644  AAGACCTACGAACTGGGCGACGCTAGCACACGCGAGCTGGGCGCGGAGATCACGCTG 4585
QY      562  TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db      4584  TTC-----
QY      582  LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
Db      4581  ---TACATCGACTTCGACGACGAGCTGACGTACGTCAGC---AACGACGTCGCTGGACC 4528
QY      602  PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyVal 621
Db      4527  AACCTCGCGCGC-----ACCAAGCACCCAGGCATCGAGACT 4492
QY      622  SerGlyAspTyr-----ValArgGly 628
Db      4491  TCCGGCCACTATGACTTCGCGCTCTCGATCCGCGCTGACGCGCTGTGCTACGCGC 4432
QY      629  ArgLeuLys-----AsnLeuProSerLeuProGlyArgGlu 640
Db      4431  AGCTCTACCTATACCCGCGCACCTACGAAAGGCGACATTCCTCTTCAAGGGCGCGAC 4372
QY      641  AsplalaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
Db      4371  CTGCGCGTCTACTCGCGC-----CAGTGGCC 4345
QY      661  AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
Db      4344  ACCCGCGGGTGGCTACGAGTGC-----GACCGCTGACCTCAACCTC 4300
QY      681  AspTyrTyrArgValPheAlaGlnAsn-----LysLeuAlaArgTyrGluThrArgThr 698
Db      4299  GAC-----GCCTTCGCGCGTCCATCGCGCTGCGCTACGACTTCGCGCGCGAGCGTC 4250
QY      699  ProGly-----HisHisMetLeuAsnLeuGlyAlaAsnTyrArgArg 712
Db      4249  CCAGGCAACTTCCACCAACTACATCCAGCCCGCGCGCGCGCGCGCTACGCGGA 4190
QY      713  AsnThrArg-TyrGlyGluTyrAsnTyrTyrValLysAlaAsp----- 726
Db      4189  CATTCGCGGTACGTGACCTGGAATGCGCGCTGCGCTACGACTTCGCGCGCGAGCGTC 4130
QY      727  -----AsnLeuLeuAsnGlnSerValTyrAlaHisSe 737
Db      4129  GAACCTGAAGCTCGGACTCGGGGTGAAGAACCTGTTGACAAGCAGTATTTCACCCGCTC 4070
QY      737  rSer 738
Db      4069  CAGC 4066

RESULT 36
AE006093
LOCUS
DEFINITION
Pasteurella multocida PM70 section 60 of 204 of the complete
genome.
ACCESSION
AE006093 AE004439
VERSION
AE006093.1 GI:12720840
KEYWORDS
SOURCE
ORGANISM
Pasteurella multocida
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Pasteurella.
REFERENCE
1 (bases 1 to 10777)
AUTHORS
May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
Kapur,V.

```

TITLE Complete genomic sequence of *Pasteurella multocida*, Pm70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 21145866
PUBMED 11248100
REFERENCE 2 (bases 1 to 10777)
AUTHORS Zhang, Q. and Kapur, V.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
55108, USA

FEATURES
source
1. .10777
/organism="Pasteurella multocida"
/mol_type="genomic DNA"
/strain="PM70"
/db_xref="taxon:747"
complement (74. .1027)
/genes="trxB"
/notes="synonym: PM0573"
complement (74. .1027)
/genes="trxB"
/codon_start=1
/transl_table=11
/product="trxB"
/protein_id="AAK02657.1"
/db_xref="GI:12720841"
/translation="MNAMKHSOLLILGSGPAGVTAIYAARANKLPVLVTGLQGGQL
TTTETENGGDFGETPELMQHAERKFDTEIFDHINKVLSRRPFPKLYGVDVH
TSCDALLIATGASARYLGLPSEAYKRGVSACATCGGFVYRNKPAVIGGNTAVE
EALYLANIASEVHLVRRDSFRSEKILIDRLMKVVEGKILIHTRNLTDLVLDGMGV
TGVRLKDTSEATEELSLDGVFIAGHPNTDIFAPQLELNNGIYVKSGLGNATAT
SDGVFAAGSDVMDHNRQATISAGTGCMAALDAERYLDAQ"
complement (1109. .1762)
/genes="PM0574"
complement (1109. .1762)
/genes="PM0574"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02658.1"
/db_xref="GI:12720842"
/translation="WVAQAFRIQTLPTTVLFKEAQAALDAPPGILDEQALQORLALILP
KEDEIKFMALEWLEAEFEQALPLKKEAWSELSAKNSDIALIYAEVTIAMKKVSPA
DILNQIPLODRUNWQGLQAQIELLIQAADTPEIQOLQSEPAKNSPQIALKALQLH
QANKREASLLDFLRLDNLNENGEIKQQLAMULTAIGQNDPVSNNKRYRLLSLYL"
complement (2098. .2787)
/genes="PM0575"
complement (2098. .2787)
/genes="PM0575"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02659.1"
/db_xref="GI:12720843"
/translation="MKKIEQLFANNYSWALRMKEENSSYFKELADHQTSPSYLWIGSD
SRYPACDLNLGPGELFHLNRVANQVHTDNLCLSVQAVAVLWIEHIIICGHTNCG
GIKAAADODLGLINWLLHRLDIWVKHSHLGNLPPEKRAMLTKINVAEOVYNLR
SSIIKSAKNGKSLSLHGWYVDYNDGFLIDQGVATSRSELSIYSYNATARKLNED
EILKQODEBA"
2953. .5181
/genes="hemR"
/notes="synonym: PM0576"
2953. .5181
/genes="hemR"
/codon_start=1
/transl_table=11
/product="HemR"
/protein_id="AAK02660.1"
/db_xref="GI:12720844"
/translation="MDKNLMKGCVFSLVGCIGIQLASNPNDVDLELLPIVNADE
DNKLPGRSVLKXNIDQQAADNADLNLPGVNMAGGFRPGSQTLNMGMDAEDVR
VOLDGATKSPKQCGSIFTEPELLRKVTVDKGNYSPOYNGGAGTVPKTDATDF
LKENQKIGGLFKYGNNSNNQKTYSTALVLQNEQKNIDLLFLGSVNASNATYRPPKSK

ILFSKNNKSLIKVNWQITPHELTLTFSVYGIHKGWEPMAAKRDVMSRPTETEIKRY
GIDVAKRKLAVYEDQDESYSLAKRYVLPENNKWNLVSQVLSYKSTENDTRHEKVTSS
FLGLNGKSWITYSDFDTSNTSTNLIGRAEHELLEGLQWLKKNKANTLMYHKEGVKK
ADNYGYFYPMPSGRQYTHAFYLDQIKQNFLEGGIRYDHINNIGKNLAPRYN
DISAGHDYSQKYNNGSYLYGLKYDVNHYLSLFTNFKTWRAPIVDEQYETQYQASV
SATSLNLEKEMINQTRVGGIITLNLHLPQENDAFQFRITTYFNRGKNEIFKTRGVNVE
NALDVNSVCPKIIISNRYNLPGVIOGALEAYYQSYLFGGLTYSVYVKGKRTSPRN
PKSTSTWIAETPRKATATLGNIPYFYTAGRAEFVRKQDRSPLSNDKASYAL
PSSKGSLSHSVFPSSNPTKIKGNFKYVDNLNRPYPYFYLGSASTGKGVKESLTQ
QF"
complement (5220. .6026)
/genes="trpA"
/note="synonym: PM0577"
complement (5220. .6026)
/genes="trpA"
/codon_start=1
/transl_table=11
/product="trpA"
/protein_id="AAK02661.1"
/db_xref="GI:12720845"
/translation="MRRFDTLFSQLOQKQCAFVPVTLCDPNLAOSFEIIQTLVENG
ADALELGFPSDPMLDGSPVIOAANORALVAGCSDASFOLLTOVRSYPTLPISLLIC
ANLYAOGIERFYQRCAEVGDVAVLIDLPLFHREFTVMAEKYSIQQVFCPPNAGE
DTIKRVACQQTGGYTYLVSRSGVTGVENQSQADMILKTLVERLKAYAPPILOQGFISOP
QQVSDVLKIGVNGAIGSATVQIIEKYHDDPQACLAELAQFTQMKQAITT"
complement (6013. .7221)
/genes="trpB"
/note="synonym: PM0578"
complement (6013. .7221)
/genes="trpB"
/codon_start=1
/transl_table=11
/product="trpB"
/protein_id="AAK02662.1"
/db_xref="GI:12720846"
/translation="MSETLNPYGFEGGMVPEILMPVLKLEKAFVEAQODPTPKE
TFDLILKNYAGRPALTFCENLQGGSKTLYLKEEDLHGAHKNTQNLGQILLAKRW
GKTRIIAETGAGHGVAITACALACMLMPCQIYGAKKDVERQSPNVFMRMLGANVTAV
TKGSALSKDACCEAMRDWAEYHEHTYLLGTAGAPHPFTTIVREFOKIIGETKQQLL
AREGLPDVAVIAVGGSGNAIGFNDPIETSVRLIGVEPAGKATGQHGAPLGHGT
TGIVFGKMAPLQMDPQGBIESYSISAGLDFPSPVQPAHLQAGRAQYBSITDDEAL
SAFOLARHEGIIIPALSAHALAYALKIQRQPEKEQLLVNLSGRGDKDIFTVDRL
SQKGSVYAPF"
complement (7423. .8181)
/genes="ydfG"
/note="synonym: PM0579"
complement (7423. .8181)
/genes="ydfG"
/codon_start=1
/transl_table=11
/product="ydfG"
/protein_id="AAK02663.1"
/db_xref="GI:12720847"
/translation="MKKIALVTGATAGTGTACRTLIQHNVNVIGTGRRAERLAALQ
EFGADFLPLADFLCDRHATEKAIINLPAQWAVDVLVNNAGLALGLDRADKAEIDWI
QMNITNTGLVTVTRLLPQWVSRNQGHIIINLGSAGTYPPGNGVYGASKAPTKQFS
LNLADLQATVTVTRNIEPGLCGSTESNVREKGDQRAAKLYVEDVQVFTPEDIANIV
LWANOQZEHVNNRIEVMPTAQSSAPLNVKQTK"
complement (8217. .9626)
/genes="trpC"
/note="synonym: PM0580"
complement (8217. .9626)
/genes="trpC"
/codon_start=1
/transl_table=11
/product="trpC"
/protein_id="AAK02664.1"
/db_xref="GI:12720848"
/translation="MTQAFPSILQKIIQDKTWTIAEQAOQLPLASFODLERSDRDPY
QAMQGSQPSYILECKKASPSKGLIRAAFPDLTEAEVQHVATVSVLTDRKYFG
HMOAQVRQSVPTOLCKDFIILSEYQVLYRAQADAILMLSVLNDERYQTALTLA
HQLGQVLTETSENEAFARALKALPKYVNNRNLDLSVDIORVAQITQHYADQIFP
ETKIISESIGYHQQIIRDLRHVADGFLIGSSLMQHTDLNHAHVRSLLFGEHKVGLTRP

QDVRVVOGALYGLLI FVTHSKRCISLRQAEIWTQAPLRFVGVFOODSIFACHIA
TQRLHAYVQHGNEHSEHLEIROLRDELVPNCEIWRALCISSSGTSPCOLVENKNDIYI
FDSKTHQOQGGGKFTQWLPKPKMLAGGITPENIPLAALQOGLGUDNSGLE
TPYGVKSHQKVAAFAQR"

gene
complement (9663..10664)
/gene="trpD"
/note="synonym: PM0581"
complement (9663..10664)
/gene="trpD"
/codon_start=1
/transl_table=1
/product="TrpD"
/protein_id="AAK02665.1"
/db_xref="GI:12720849"

/translation="MTEQLHOLFPEKKALNOOQTEFLFTAIRGOLONSOLSAALIA
LKLGETPEEISGATLQADAEPIIDYPPADIVGTGGDANTINISTASAIVAAS
YGLKVAHGNRSVQSQTGASDLITAGLVNHHLSAHKARQALDEIGLCFLPQVHLGF
QHAIPVQALKRTIFNIGLPLINAKPKRQLGLVYSSDLQPIYAEVVARLGHESVV
VHSGLDEITAHQVTEVAIEKAGQIERVYLTTPDFQFPQPLETLURGKPSNQAQILT
ALLOGKGLAHQQAAMNSMLSLFGYPENLKONQAQIMDIATGKPFETLOOLTOY"
BASE COUNT 3178 a 2042 c 2253 g 3304 t
ORIGIN

Alignment Scores:
Pred. No.: 1.6e-05 Length: 10777
Score: 253.00 Matches: 165
Percent Similarity: 35.47% Conservative: 117
Best Local Similarity: 20.75% Mismatches: 291
Query Match: 6.27% Indels: 222
DB: 1 Gaps: 39

US-09-936-377-2 (1-758) x AE006093 (1-10777)

QY 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db 3133 CAACAAACAGCAGATAATCGTGTGACTTAATAATAATTTTGTGCTGTAATAATATGGG 3192
QY 87 GlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgGlyLys 106
Db 3193 -----GGAGGATTTCGCCCT-----GGTGGTCAACA 3219
QY 107 ValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 3220 ----TTAAATTAATGA---ATGGGTGATGCTGAAGATCTTAGAGTTCACTAGACGGC 3273
QY 122 -----ProAspHisAlaIleMetValAspThrAlaLeuSer 133
Db 3274 GCAACAAAAAGTTTCGAAAAAATATCAACAGGCTCTATTATTTATGAACCTGAGTTATTA 3333
QY 134 GlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAla 153
Db 3334 AGAAGGTGACAGTACACAAAGGAATATTCTCTCAATATGGCAATGGTGTCTTGGCT 3393
QY 154 GlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn---GlyVal 172
Db 3394 GGTACTGTAAATTTTGAACAAAAAGATGCAACTGATTTTGTGAAGAAAAATCAGAAAAATA 3453
QY 173 SerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGly 192
Db 3454 GTGGATATTATTAATATGAATAATATAGCAATATATACCAAGAACTTATAGTACAGCC 3513
QY 193 IleAsnIle---GlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys 211
Db 3514 CTAGTTTTACAGATGAACAAAAAATATTGATTTGTTATTATTATTTGTTCTGTAAGAAAT 3573
QY 212 SerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArg 231
Db 3574 GCAAGCAATTATCAAGACCT-----GATAAAAGTAAATTTCTTTTCAAAAAAC 3624
QY 232 ArgPheAlaAsnGly-----GlnHisArgAlaVal 241
Db 3625 AATCAAAAAAGTGGATTATAAAGTAATTTGCAAAATTAATCTCTGAAATTTATTAATCT 3684

QY 242 Leu-----GlyTyrArgLysArgPheTyrArgThrTyr 253
Db 3685 TTATTCAGTGTATGCAATTCATAAGGGTGG-----GAACCTTGG 3726
QY 254 SerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHis 273
Db 3727 GCAGCAAAAAGAGATGTGATGTCGAGACCAACAGAAACA---GAAATPAAAAACGCTATGGG 3783
QY 274 AlaAspIleLeuTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHis 293
Db 3784 ATTGATGTTGGTGGAAACGTAACCTTGTATTATCGA----- 3819
QY 294 LeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAsp 313
Db 3820 -----GATCAAAAAGATGAAAGTATTATTCATTG----- 3846
QY 314 AspAspAlaHisAlaHisAlaHisAsnGlyLysProThrIleAspLeuArgAsnLys 333
Db 3847 -----AAATATCGCTATTACCTGAAATAATAGTGGATTAATTTATCTGTTGAG 3897
QY 334 ArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgVal 353
Db 3898 CTGAGTTATATGTAACACAGAA----- 3918
QY 354 HisLeuAsnArgAsnAspTyrHisHisAspGluLysAla----- 366
Db 3919 -----CAGAATGATCTACTCGCCATGAGAAAGTCTCTTCACTCTAGGTACATTA 3969
QY 367 -----GlyAspAlaValGluAsnPhePheAsnAsnGlnThr 378
Db 3970 GGAATAAAGTTGGATACTTATTCAGATCTTATTTGATATAGTAACACAGTACT 4029
QY 379 GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer----- 396
Db 4030 CTAAAT-----ATTGGCGTGTCTGAGCATGAATCTACTA 4062
QY 397 TrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVal 416
Db 4063 TTTGTTTACAGTGGTTAAAAAATAAAGAAATACCTCTATGTATCATAAAGAGGGAGTC 4122
QY 417 Lys-----GlnProMetLeuLeuAspAsnLysValGln 427
Db 4123 AAGAAGGCAGACTATAATTATGGCTTATTTTTCAGCTTATATATGCTCTCTGGACGCCAG 4182
QY 428 His---TyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGly 446
Db 4183 TATACACATGCATTTTATTTTACAGATCAATAAATGGCAGAAATTTCTCTTTTACAGGA 4242
QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAsp----- 460
Db 4243 GGG-----ATNAGATATGACCATATCAATAATATAGGG 4275
QY 461 ---LysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAla 479
Db 4276 CAGAAAAATTTAGCGCCACAGATATAAT-----GATATCTCTGCA 4314
QY 480 HisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyr-----Phe 495
Db 4315 GCACATGATTATAGCCGAAAAATTAATGTTGTTCTTATTTATTTAGTCTTAAAGTAT 4374
QY 496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu-----ArgLeuProSer 513
Db 4375 GATGTAAATCATTTTAAAGTTTATTTACGAAATTTTAGTAAACCTTGGCGAGCACCTGTT 4434
QY 514 ThrGlnGluLeuTyrAla-----HisGlyLysHisValAlaThrAsnThrPheGlu 530
Db 4435 ATTGATGAACAGTATGAGACACAATATAGTCAAGCTTCTGTATCTGGACACTCTTGAAT 4494
QY 531 ValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyr 550
Db 4495 TTAGAAAAAGAAATGATTATTAACACACAGCTGGTGGAAAT---ATTACTCTCAATCAT 4551
QY 551 -----GluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567

FEATURES		Location/Qualifiers	
source			
Db	4552	CTATTCAGAAATGATGCTTTTCATATTAGAACTACTATTATTTTCAATCGCGCAAG	1..10399 /organism="Plesiomonas shigelloides" /mol_type="genomic DNA" /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	568	AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProlyssSerIleGluAspAsp	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4612	AATGAATCTTCAAAACG-----AGAGGGTTAATTTCGCTAGAAAATGCT	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	588	SerGluMet-----LysLeuValArg-----TyrAsnGlnSerGly	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4657	TTAGATGTTGATATAGTGTGTCCTAAATATTATTAGTAATTACGTAATTACCTGGT	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	600	AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheIysProThrProArgTyrArgIle	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4717	TATGTTATCAAGGAGCGAATTAGAAGCTTATTATCAATCATCGTATTATTGCT---	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	620	GlyValSerGlyAspTyrValArgGlyArgLeuIysAsnLeuProSerLeuProGlyArg	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4774	GGATCGATCATATCTTATGTAAAGAAACCGGATCTTCCACCAAGAAATCCATGGAGT	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	640	GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4834	AAACATCT-----ACATGATCGCA-----GAAACACCATAGAAAA	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	660	ProAlaAlaArgLeuGlyPheHisLeuIysAlaSerLeuThrAspArgIleAspAlaAsn	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4873	GCAACCGCTACTTTAGGTTTAAATATTCGGAATAT-----	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	680	LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4909	-----TATTTACGGCAGGTGCGTGCTGCTGAGTTTGTGAGAAGCAAGATACATCA	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	700	-----GlyHisMetLeuAsnLeuGlyAlaAsnTyrArgArg	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	4963	CTATCTAATGATCTTAAAGCATCATATTGGGCATTACCTTTCTTCAAAAGGATATAGC	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	713	AsnThrArgTyrGlyGluTyr-----AsnTyrTyrValLys	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	5023	CATAGTGTATCTTCTCTGGAGTCTTCAAGATTAAGAAATGAATTTCAAGTTACT	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	725	AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerPheLeuSerAspThrPro	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	5083	GTTGATATATTTATTAACCGACCTATAT-----CCT	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Qy	745	GlnMetGlyArgSerPheThrGly---GlyValAsnValLysPhe	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
Db	5116	TACTTAGGAATAGTCTTCAGAACAGTAGGAGTATGTCAAATTT	10399 /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
RESULT 37	AY008342/c	10399 bp DNA linear BCT 23-APR-2001	
LOCUS	Plesiomonas shigelloides heme iron utilization locus, complete		
DEFINITION	sequence.		
ACCESSION	AY008342		
VERSION	AY008342.1	GI:13774058	
KEYWORDS	Plesiomonas shigelloides		
SOURCE	Plesiomonas shigelloides		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
REFERENCE	1 (bases 1 to 10399)		
AUTHORS	Henderson,D.P., Wyckoff,E.B., Rashidi,C.E., Verlei,H. and Oldham,A.L.		
TITLE	Characterization of the Plesiomonas shigelloides genes encoding the heme iron utilization system		
JOURNAL	J. Bacteriol. 183 (9), 2715-2723 (2001)		
MEDLINE	21189235		
PUBMED	11292789		
REFERENCE	2 (bases 1 to 10399)		
AUTHORS	Henderson,D.P., Wyckoff,E.B. and Payne,S.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-SEP-2000) Science and Mathematics, University of Texas of the Permian Basin, E. University Blvd, Odessa, TX 79762, USA		

1..10399	/organism="Plesiomonas shigelloides"
/mol_type="genomic DNA"	
/db_xref="taxon:703"	
complement (386..943)	
/gene="hug2"	
complement (386..943)	
/gene="hug2"	
/note="heme utilization protein"	
/codon_start=1	
/transl_table=11	
/product="Hug2"	
/protein_id="AAG23392.1"	
/db_xref="GI:13774053"	
/translation="MKSEDIYNTNERQBELQNLQPEIAEPNECRTLVLATVDADGMP	
RVSTAFVQLTDGTYVLISETARHARNLQQVPTSLMMIEDSESQRLFAKRLTFLDA	
NYEKVARDTAQWAAVALEAHGDIIVANLSQMAADFILFRWPEQGLYVKGQAFQV	
SSDLVDVHLQEGHKTSGEPVKA"	
complement (1108..1590)	
/gene="hugX"	
complement (1108..1590)	
/gene="hugX"	
/note="heme utilization protein"	
/codon_start=1	
/transl_table=11	
/product="HugX"	
/protein_id="AAG23393.1"	
/db_xref="GI:13774060"	
/translation="MSASTQBLAARFGVSEYQVLAALPEAMVMAHGSMAQSLLAEELT	
DWPLTTTISESGEYIFKPGAPFVKGPGHGFNLYSREGTGLHGLHLLSQVSHIALLA	
KPFMGKASYSLOFFAANGRTWFKYGLGNKORQLPEQVALFEQLRAKRLIAQPOPAAE	
"	
complement (1814..3169)	
/gene="hugW"	
complement (1814..3169)	
/gene="hugW"	
/note="heme utilization protein"	
/codon_start=1	
/transl_table=11	
/product="HugW"	
/protein_id="AAG23394.1"	
/db_xref="GI:13774061"	
/translation="MMIDLSVLDLSILGHSPDPLAVATGKQAHAGSGRPLPPP	
AVQQLQMLQMDTPAIRPCLYIHIFPCVRKTCFCSFYAASSTLVAFAALUSEE	
LKRAAMPWTQAAPFAAYIIGGTEPTDLTAEQIFALQILKAHPELTDCCITLEGRL	
NRFSDEKYHAAVDGVRNFRSFGVQSDTDVRRKARLDREKLETLVHVALDRAPI	
IVDLMFLPHQSATAWQDLQDFLHSGAHGVDLYOLIEMOQTPMARLVENRLQAPAA	
TPEKSRMVAQGVAFMCOHQIQQISVNHWTRDRERSYNHLAKTTADVLPAAGAGGF	
IDGFALMOTRDLSEYHOAIRERYEVAMMSPNPOHALHCAIKAGFDRGLQRSLEDS	
LSCMALYIVSVPLLTQWQERGLLSLEADQARLTAGSFNINMAQGLINLRSLPPLA	
A"	
complement (3169..5235)	
/gene="hugA"	
complement (3169..5235)	
/gene="hugA"	
/note="heme receptor"	
/codon_start=1	
/transl_table=11	
/product="HugA"	
/protein_id="AAG23395.1"	
/db_xref="GI:13774062"	
/translation="MPVQSGAIFRSTVSATAAMTAKHTGFKRSALSLILNSVAVSF	
ANAAAPSPOTQNEVLVATNTRENKALSQETRSVAVVTQAQIQQOQPSVAELKYIP	
NYVSGGFRANQSPSVGLSGNRILQVVDGVRQNTTSGHRLNSWDEMLSSLEVIK	
GPTSLWGSGLGVAQQTISAPDMLHGDQSVGGYLRQGVASANDETKTSGLVTVGLL	
GDSVDFLLNGYYADANRLNGEHLTDSAYNRGGLAKFGWQADQLRLSVRQAE	
YRVTKQSDKVDYRTLGNISHSQFDLLSLTYGGDIYEDKTKQGREKNRPIDAGR	
SKVMSGYQADIPLGSONWLLPGLYDHFETAEKNIAGSERSEHLSFSVGLRWAAT	
WLTILNARDEAFAPSMEMTYTGHFCMGPGCNVFNANFDLPETAKNEISNQMR	
FNWVLADLAFGATYFNNVSNYIDQQNMFTQYVNVTDARLGLVGLDARLNRDL	
ETSLSYAQTEGDKTKNQPLNNIIPAKHWLVGLVSHYFMDRLKAGVNVSHYSQDELPS	

183 GlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLysAsnPheVal 202
 4626 GCCAATGATGAACCAACCAACGCGGTGCGTATGCGCTGTTGGGGGATTCGCGCAT 4567
 203 LeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsn 222
 4566 TTCCTGCTAACCGCTACTACGCGGATGCC-----ATAAC 4531
 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
 4530 CTG-----CGTTGGCAATGGCGAGCAC----- 4507
 243 GlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeu 262
 4506 -----CTGACTGACTCGGCCCTATCGCAATCAGGCGGGTTA 4471
 263 ProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeu 282
 4470 GCCAAATTTGGCTGGCAGCGAGATGACGCG-----CAGCGTCTGGAG 4429
 283 IleAsnLysArgTyrLeuGlnLeuTyrProHis----- 293
 4428 CTGTCAGTTCCGCAACCGCAACCCATCAGAACCGACCGAGCAACCCATCAGAGATGGTC 4369
 294 -----LeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSer 308
 4368 AGCACCTCAGTGGCGCTGGTGCAGCAGAAACCGTGATTTCC--AATACCACTTGGAT 4312
 309 CysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrpIle 328
 4311 TATCGCTCATCCCAACCGATAGCCGT-----TGGTTG 4279
 329 AspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPhe 348
 4278 GACCGACGCGTAACCTGCTATGTGTGAACAAACGAA-----TTT 4240
 349 GluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAsp 368
 4239 GATGAATATCGCTGACGAAAGCAAGCAAGGATAAGGTGATTACCGCACACTGGGTCTG 4180
 369 AlaValGluAsn-----PhePhe 374
 4179 AATATCAGTACCACTGCGAGTTGATCTGCTGAGCTGACCTATGTCGCGCGATATTTAT 4120
 375 AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle-----Gly 391
 4119 GAAGATAAAACCCAAAGCGAGCGG---GAAGGGAAGAAACCGCCGATTCACCGCGATGGT 4063
 392 ArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla 411
 4062 GCGAGCAAA---GTCTGGGA-----AGTTATGACAGGCC 4030
 412 ThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhe 431
 4029 GATATTCATTAGGTAGTCACTGCAACCTGTTACCGGGT----- 3991
 432 PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGlu 451
 3990 -----CTGCGTTATGACCACTTTACCGCGAA----- 3964
 452 LysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys 471
 3963 -----GACAAAGAATATCGCGCGCTCCGAGCGTAGCGAGAT 3928
 472 GlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGly 491
 3927 CACCTGTGCGCATCCGTTGGCGCTG----- 3904
 492 AsnTrpTyrPheTrpGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu----- 509
 3903 CGTTGGCGGCGGACATGATGG-----CTGACCTTAATGCTCCCTTACGAGCAAGCGCTTC 3850

510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis-----Val 524
 3849 CGTCACCGAGCATGGAGAGATGTACACCAACCGCTACTCTCTGATCGGCCCGCATG 3790
 525 AlaThrAsnThrPheGluValGlyLeuAsnLysHisLeuAsnLysGluArgSerAsnAsnIle 544
 3789 GGCTGCAACGTTTAA---GCCAATCCCGATTGAACCCGAGAGCCGCAAAACAAA 3733
 545 GluLeuAlaLeuGlyTyrGlu-----GlyAspArgTrpGlnTyrAsn 558
 3732 GAGATCTCAGCTCAGATCGCGCTTTAACAAACGTCGTCAGATGATGATAGCGTTTGGG 3673
 559 LeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln----- 573
 3672 GCGACTTACTTCCATAACACGTCAGCACTACTTACACGAGCAGCAACATGTTCCACC 3613
 574 -----ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu 589
 3612 ACCCAATACGTGAACGTCAGCGATCGCGT---TTGCGGGCGCTCGAGCTGATGCG--- 3559
 590 MetLysLeuValArgTyrAsn-----GlnSerGlyAlaAspPheTyrGlyAla 605
 3558 -----CGTTACTTGGCGCGATCTGGAACACGCTGAGCTATGCCAGACA 3511
 606 GluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyr 625
 3510 GAAAGG----- 3505
 626 ValArgGlyArgLeuLysAsnLeuPro-----SerLeuProGlyArgGluAspAlaTyr 643
 3504 ---CGCGACAGAAACCAACGCGCTGATACATTCGCGCATTAAGTGGGTATTG 3448
 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
 3447 GCGCGTGGCCACTACTTATGATCGCGAC----- 3418
 664 LeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyr 683
 3417 -----CTGAAAGCGGCGTG-----AACGTGACGCACTTAC 3388
 684 ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArg-----ThrPr 699
 3387 GAA-----TCACAGATGAATGCCATCTAATGTGACCAATGTATCCGCTCTACACC 3334
 699 OGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTyr 719
 3333 TTGTTGGATCTCTACACCACTTGGCAGCCACAAACGCGCGCTCAAAGCAGTCAAAGTG 3274
 719 pAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPh 739
 3273 GATGTGGGT-ATC-----GATACGTGACTGACGAA-----TATTACCGTCAGCGCTT 3227
 739 eLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 3226 T---GACCAGCTCTATTCCGCGCTGTAACCTTAAGTGTGTCGCTATTCCTTC 3172

RESULT 38

AE015838
 LOCUS
 DEFINITION
 Shewanella oneidensis MR-1 section 387 of 457 of the complete genome.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AE015838
 AE015838.1
 GI:24350433
 Shewanella oneidensis MR-1
 Shewanella oneidensis MR-1
 Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 1 (bases 1 to 12558)
 Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W.,
 Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R.,
 Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S.,
 DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R.,

Peterson, J., Umayam, L., White, O., Wolf, A., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*
Nat. Biotechnol. 20 (11), 1118-1123 (2002)
22297686
12368813
2 (bases 1 to 12558)
Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W., Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Direct Submission
Submitted (12-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
Location/Qualifiers
1. 12558
/organism="Shewanella oneidensis MR-1"
/mol_type="genomic DNA"
/strain="MR-1"
/db_xref="taxon:211586"
complement(183..335)
/locus_tag="SO4065"
complement(183..335)
/locus_tag="SO4065"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057039.1"
/db_xref="GI:24350434"
/translation="MCLNVRCLATFLLITNSLLSLRKLKVKKNSGFGVDNKGGRDGG"
389..1492
/locus_tag="SO4066"
389..1492
/locus_tag="SO4066"
/note="identified by match to PFAM protein family HMM PF01259"
/codon_start=1
/transl_table=11
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase, putative"
/protein_id="AA057040.1"
/db_xref="GI:24350435"
/translation="MSLADSVLAVNNDLPRTDKSVHSGKVRVYVWLTADSRLLIKT KGNVPDETLAVISDRISAPDCIFHGGGLKIPGKGAALNAINSHWFKLFAENG LADSHLDIPFPFWVQKARPKEVAKCHQYITGSMWYRSGKGVFCGIIPEGLE KDQPELLITPFGILTGIPGPAQDDNIRSDIEANYQAPGEKVEDIDLYEKL RDGFKVISKALADLDQVVDVTFEFGYVTDQDNGSKLYMDEVGTPDSRSDWGAAY RDGKILNKGSRFQFLNHPFDPDILLNDRMPERBALDARNALPLEAMMQVSRITYT GIAEKVTAAPLPAKPAKIIKILREYDLIV"
complement(1828..1974)
/locus_tag="SO4067"
complement(1828..1974)
/locus_tag="SO4067"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057041.1"
/db_xref="GI:24350436"
/translation="MLNFKVWFGITLFAHYSQATPKPKQYTFPCPEVAGRWKRLRE SGHG"
2140..4248
/locus_tag="SO4068"
2140..4248

/locus_tag="SO4068"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057042.1"
/db_xref="GI:24350437"
/translation="NNKRDIEHLLSLAVFSOTYKNNTDIDVIAAFIETATIKRALK SFTCPBLTARVNSEFTPEGVISTSLAKISHLEATKPKTHYIYTPDQOESNKS MLLEISNAEKAKALLVDLEGYISSKIEKSESEVDDLIHSFTSFIQDTANGKPFSE HISAFTIENSEKADPIEHVQOEIKGDSIISEIRFNINISSETSGSWNEITILLGTILIF HLAYNGLVFKITIDFELSYISEINSGRKEIKLRYFNELTLOGINSYFKADDIYAK DKALPHVAMVLTNGCESRFDLLKKKELIRLEVKIKPLEDSFFSEENHKYNI1 NSGIIINEOSDPNLMENHINELYNILIRKGDSSRFDRITHIIVKYNITNIAWS SSILKEGVEPLATNFMISKWFLNRGFGOGSITSPDVIKAKIVLSSKVNESVS KRYTEQKVSNTLDEKATATLAECRSIDTRKPEIKAVETOMLSLIDESSIDAYL RKNHLEQINSRESNAKMLKLVHERENDINIKTLIESKLTTKLYQEKQSTTQ QLTFFESKRTDINKTIRSKTKLILNTLITVIPSIIYLLSEDEIKETNLIKISGA TTISTFIWYFGDMIKILNRTISKFDNHLNKKFEKKGLTDEIILKQUNDEISKLUN DISNIDQDIKILELEEQN"
complement(4733..5518)
/locus_tag="SO4069"
complement(4733..5518)
/locus_tag="SO4069"
/note="identified by match to PFAM protein family HMM PF01209"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA057043.1"
/db_xref="GI:24350438"
/translation="MSSNALYTDLSGYDLMCSIDINYQOOSASIHRLHOFGLNNGKOH LDLAGCGPHVRHFDILGYQSGLDINQPMLDNMACRPERQNLQMDTAFYLEQPYD LITCFYLSIHSAGIEIKSLASVHRALMDGGIFCFNSVDKHCINRSLUSIKHFAFEE YAFAPRSWHYNGQKQKSLNSIAKQQPFTQLTDFEADTSLHQTETLMDQDHTWA TSFAELQALLAPYFEVHIFRHYEKIYVPDLSGNALFVCVKI"
complement(5783..6496)
/locus_tag="SO4070"
complement(5783..6496)
/locus_tag="SO4070"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA057044.1"
/db_xref="GI:24350439"
/translation="MKHSLALVAVGLFGSAGTATTLAAEVDPPIHETWGVQIQVE PDMAELNVEAVTEKDAKADNSDAAVAKFIARLAAAGSKHIGQANLNLOPQYII AQDKAPELTGYIASRQITVTVELSLRNLITVDSALIEGMARVNNIALSKAEAYVAK ARQBAIVDAKQKSLAKGFGENLKIWIQYRYDQRPVQVPMLEMMNAKADYAMAAES YQYGVQVITEDRVEVYVYK"
complement(6671..6796)
/locus_tag="SO4071"
complement(6671..6796)
/locus_tag="SO4071"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057045.1"
/db_xref="GI:24350440"
/translation="WYSQDGVSTSYSPHNAVFRYQSAFRPAGIHFRARSVQGRF"
6888..8333
/gene="y11g"
/locus_tag="SO4072"
6888..8333
/gene="y11g"
/locus_tag="SO4072"
/note="identified by match to TIGR protein family HMM TIGR01579"
/codon_start=1
/transl_table=11
/product="MiaB-like putative RNA modifying enzyme y11g"


```

Qy 433 GlyValGluGlnAlaAsnTrpAspAsnPhetThrLeuGly-----GlyValArgVal 450
Db 11553 -----AATGCAATGATGGCATGCCCAATGTGGATGCC 11588

Qy 451 GluLysGlnLysAlaSerIleArgTyrAspLysAla-----LeuIleAspArg 466
Db 11589 GAGCGCGCAAAAGCCTTATATGCGATGCACAGAGCGGACGATACCTTAATCGATGCG 11648

Qy 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSer 486
Db 11648 ----- 11648

Qy 487 PheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSer 506
Db 11649 --ACATTGCTGGCAGCGTCAATTTGTCGCCCAACAACAACCTGGAATTTGGCTTAGCC 11705

Qy 507 HisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThr 526
Db 11706 CGTAAATAATCGGCCCTTAACCTTATGAGCGTTACAGTTGGGCGCGCGGCTCATGCCA 11765

Qy 527 AsnThrPhe-----GluValGlyAsnLysHisLeuAsn 537
Db 11766 ACCACTATGTTGGCTGATGGCAGCGTAACGTTATGTCGCCAACCCCGATCTAAAG 11825

Qy 538 LysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTfPglTyr 557
Db 11826 CCGGAACCGGCCACACCTTAAGTCAGCTTACAAATTTAATGGCGATGCGATGCGATTC 11885

Qy 558 AsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp 577
Db 11886 TCGACCACGGCTTGATAGCGAGTACCGATATATCGATGCTGGAAGTGATT----- 11939

Qy 578 GlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGln 597
Db 11940 GCGACCTTTAACCGTACCAGTACACCCCGCAGTAAACGCAATATCTCTAAATTCACCAAC 11999

Qy 598 SerGlyAlaAspPheTyrGlyValGluGlyGluIleTyrPheLysProThrProArgTyr 617
Db 12000 GAAGATGCCAATTTATATGGCGCTTAAGTTACGGCAGCTCTAC----- 12041

Qy 618 ArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeu--- 636
Db 12042 -----CTATTAGCGGATACCGACAGTGGTAAGTGGCAGATGCTCGGTAAACTCAAC 12092

Qy 637 -----ProGlyArgGluAspAlaTyrGlyAsnArgPro-PheIleAla-GlnAspAspG 654
Db 12093 ATCACCGTGGCGAGCGGATGAA---GGTAATGAACCACTTTTCAATCAAAACCGTTA 12149

Qy 654 InAsn-----AlaProArgValProAlaAlaArgLeuGly-PheHisLeuLys----- 669
Db 12150 CAAACCGAGTTAGCCCTCAGCCACCAACTCGCGGATTCGGAACCCGCTTGGCATGGCAA 12209

Qy 670 ---AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGln 688
Db 12210 TGGGTGGCGACAAAGATCGGCTTATGATCGC----- 12242

Qy 689 AsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAla 708
Db 12243 -----CGTTAGAGAACCAACGACGACACTATTGCTACTGAACTTAAGCAGT 12290

Qy 709 AsnTyrArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeu 728
Db 12291 -----AGCATCAATGGCAAGAGCTGAGTTTAACTTTCGATCAATAACCTG 12338

Qy 729 LeuAsnGlnSerValTyrAla----- 735
Db 12339 TTGTAT---ACTTACTAGCTGCCACTCGTGGTGTGAAGCATTCGCGATTTAAGGCC 12395

Qy 736 ---HisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyVal 754
Db 12396 GATAGCAGCAATGGCTTTAGCCAAGTCGCGGCTCGGCTCGCTCATTTGAATTAGTGCA 12455

Qy 755 AsnValLysPhe 758

```

Db 12456 AGCTACCGCTTC 12467

RESULT 39
AE000694/c
LOCUS
DEFINITION Aquifex aeolicus VF5 section 26 of 109 of the complete genome.
ACCESSION AE000694 AE000657
VERSION AE000694.1 GI:2983162
KEYWORDS
SOURCE Aquifex aeolicus VF5
ORGANISM Aquifex aeolicus VF5
Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
REFERENCE 1 [bases 1 to 16256]
AUTHORS Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
The complete genome of the hyperthermophilic bacterium Aquifex aeolicus
JOURNAL Nature 392 (6674), 353-358 (1998)
MEDLINE 98196666
PUBMED 9537320
REFERENCE 2 [bases 1 to 16256]
AUTHORS Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olson,G.J. and Swanson,R.V.
Direct Submission
JOURNAL Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121
COMMENT Putative indicates no similarity to known proteins
Hypothetical indicates similarity to a protein of unknown function.
FEATURES
source
1..16256
/organism="Aquifex aeolicus VF5"
/mol_type="genomic DNA"
/strain="VF5"
/db_xref="taxon:224324"
63..476
/gene="spsK"
/note="synonym: aq_518"
63..476
/gene="spsK"
/codon_start=1
/transl_table=11
/product="spore coat polysaccharide biosynthesis protein SpK"
/protein_id="AAC06769.1"
/db_xref="GI:2983168"
/translation="MNGKAGASSKGGNFVYTIIRKAKAGEPLRVVDDIYMSPTYTLDA
AKEIKLLENKPYGIYHVTNSGYCWSYFAVKILEYSGLKTDIKPVRHTEFKTANR
PLWSPLASRGKIKLRNWEALKDFINALINLSY"
445..1518
/gene="rfe"
/note="synonym: aq_519"
445..1518
/gene="rfe"
/codon_start=1
/transl_table=11
/product="undecaprenyl-phosphate-alpha-N-acetylglucosaminyltransferase"
/protein_id="AAC06768.1"
/db_xref="GI:2983167"
/translation="MLFPLITLISFPLCFLEIKYINLNFHDQKGVQKFFHRRTPR
IGLAIYSLVLLSVIFILNKPFSKERFLFMSAPPVFPFGLLEDITKXSPKRWLL
IGFISGLAPYLISAKVTVDIPFVDSILSILSLTAFALAGVSHAFNIDGPNP
LASGVAMLVGAYAVSVFLHNDLFVLNLLIATLGLFLLPPLIFPGDGGAYF
LGFCAGAILVKNYPDIPWPMMLLVIPVWETLFSIYRRKFLKDYPPHMPDALHF
HOLLXRLKFTIGSNIEPFKRNSTSPFLWMLLICLVPAVLFDWNTPLILMPTFAF
IVFYTWLYFRIVKFTPKIFRIK"
1519..2412
/gene="murB1"
/note="synonym: aq_520"
1519..2412

```

/gene="murBi"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylenolpyruvoylglucosamine reductase"
/protein_id="AAC06766.1"
/db_xref="GI:2983165"
/translation="MLFLKNVPLQNLTIKIGGRVSFYAEPSDLKEISLCLDFSKSBD
IFLVLGNSNTIFGVDGLVNLKAGFKVEIKGFFVEAFSGPLKDLIRFSYK
ENKGFYKLLGPAAGVSNMAGAFGEISDFLKEVYFVDEGKQLQAKRDELNES
YKSPPKLGVKVFYEFERSENILPKYKIRIRKEQPINLPTSGSTFQNPEN
FAGKLEKAGLKGFRKLVNKFSEKHANFLVNGGTFSEVVDLINIAKERYVENFGIV
LREEVKLISSGSDGKWVLA"
2369..3244
/gene="ddla"
/notes="synonym: aq_521"
2369..3244
/gene="ddla"
/codon_start=1
/transl_table=11
/product="D-alanine:D-alanine ligase"
/protein_id="AAC06765.1"
/db_xref="GI:2983164"
/translation="MKRVVMGGRSREISLTKGKAVAKALRELCHEVYELDLKEL
PCLLEIKPKDVFIAHGRYGEDTVQGLLEILDIPYGSUDTIASAVSIDKFTKIV
KSLGINTPMETFISEGVDNTWKNPFAVPREGSVGLKIVESLEELKEAYOL
LKTERVMVEEFVEGRDMTVGILKEALPVEIIPKGVYDYCEKYTKMGSEYRILKO
EKLKKLOISLKSILKOPARIDFRVTKGKIIFLEVNTPGMEITELSLPMAAK
EKMDPKLISIIIT"
3255..3905
/gene="aq_522"
3255..3905
/gene="aq_522"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC06775.1"
/db_xref="GI:2983174"
/translation="MRGILGLVPSVILLIYLSFLMKHPFQIREIKVSGLYVEIE
KIEKDHALGRLGLVLPESNILEFNEKLNRPQGVENKRFSTEGITTEMFQRKA
ISFVNEKFLMDGEGIFPWEYQKPDFTFYIYKEVFPFQKILKLLTQGNISVKY
YKDKVMVDIGGRFILLPLEDINEKEIHLMKLLKLHPDAKIFDIYKGFILLNEG"
3905..5155
/gene="ftsA"
/notes="synonym: aq_523"
3905..5155
/gene="ftsA"
/codon_start=1
/transl_table=11
/product="cell division protein FtsA"
/protein_id="AAC06770.1"
/db_xref="GI:2983169"
/translation="MKRLAALDIGSKTVFVIGERDSYGDIIHIGFVPSRGIVKGV
INDLSAKGSILRAMKEAMAGLVREVVYVNGGTTKNGTVKSNKDTISIKVS
SELESHIQRLHERCLMAKEGEYIVTAPKXILDDHTEVKNPLGLVSKLSVEMH
VVKVSTILRLKAKINEVGNPVRVNAIASDVLTHDEKEDGVLLDLMGAGLTD
YSYTEGRPIYTGVPFGNITDLSYMLKIDTAEVSKVNHGVAFESLYDDDDV
KIRPREDRIPYQKQVABEIIQSRVEEIVKFEIKARGVPLQINSIGVYVGTGA
NLKIRLEIHHMTGLPVRIQGLVQIGVLRKEIKENPKYATVCGILLRQAFVSGTIDITES
TNSGTNFFKNLSGRFKKWPPELL"
5168..6271
/gene="ftsZ"
/notes="synonym: aq_525"
5168..6271
/gene="ftsZ"
/codon_start=1
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AAC06771.1"
/db_xref="GI:2983170"
/translation="WEEFVNPCKIKVIGVGGGSGNVMRYEDGIEGVELYAINTDVQ
HLSTLKNPKIQIGEKYTRGLGAGKEVGEAALEDIDIKILELDDTMVFTISAGLG
GGTGTAAPYIAKTAKEMGLTVAVALTPFRFEGPRMEKALKLEKESDAXIVI
HNDKIXELSNRTLTIKDAFKEVDVLSKAVRGITSIVVTPAVINVDFAVRTLBEQG

```

```

LSIIGMGRGDEKADIAVEKAVTSPLEGGNTIEGARLLVITWTSEDIYDVIDDEVV
ERHSKVPFASIIIFCAVLPEQSQDFIRVAIVATDFPEKQVGEKVKFKVKEEK
EPKPEPLSDTITYLEEIPAVIRRNKKLL"
6268..7032
/gene="moac"
/notes="synonym: aq_527"
6268..7032
/gene="moac"
/codon_start=1
/transl_table=11
/product="molybdenum cofactor biosynthesis moaC"
/protein_id="AAC06772.1"
/db_xref="GI:2983171"
/translation="MRTVDITTKIREAKAYGIRLKPETVKLIKENKVPKGNLVE
ATKLSIGFCAKKTGELLPCHPIDLDVALEKVNEDLEVSTVGRGIARTGYEAL
TAVTALLNYDMCKALDDSMVIERVKLESGKSDMFRDLGVKVLNHAENEGLRK
IADYDKELGATFAEABHYISIGNLPIKNEIRSLERVISLYDFRRNPKEVKEIRV
GMSDDALIILPESEKIRFFPFTFGGIIGNLLCRR"
7020..7589
/gene="dedf"
/notes="synonym: aq_528"
7020..7589
/gene="dedf"
/codon_start=1
/transl_table=11
/product="phenylacrylic acid decarboxylase"
/protein_id="AAC06774.1"
/db_xref="GI:2983173"
/translation="MOKIALCITGASGVYIGIKLVLELDFSDVLVTSRNVKVLK
EHSLTPEVSLKLNVRIDERNFTSPLAGSRVHYGVVVPVPCSTNTLSIANGI
KNLHHRVGEVALKERVPLVLLVREAPYNEIHLNMLKITRMGVVVPASPAPFYHKPQ
SIDDMINVVVKLLDVLRIEHLNLYKRWG"
7020..7589
/gene="oprC"
/notes="synonym: aq_529"
7020..7589
/gene="oprC"
/codon_start=1
/transl_table=11
/product="outer membrane protein c"
/protein_id="AAC06764.1"
/db_xref="GI:2983163"
/translation="MKRTLGLAFSAFAQVNLBEEIOVIGKREVLTEDEVRELPAK
DVGEALYNIIPGVKVKGAJANDVYIRGPKDEVNOLFAGARVYNACNRMDPGIFH
VDFSEVSEVUIKGFEDVNYGAVGTAVTKKEPKMEGRITTVADNWSNTINPSFY
FSTGKRLSEFLIYAFRKPVEDGKRIKTELYPSGNFSAISYDEPSTANIHTAW
AKLYKITDGVKFKLDYRQATDVLIPYIMMDGIYDVRVNLGLEGRKFAQLYGS
SVRHWMTNPKRVISQNAKPYNGTYAKSKVYKGFSGFSGFIDIFIRYKAAQI
TMYRNMGVYRTQNTIPDVVYPSGNVQVYELKEGIELFAGLGSARVPPQERYFALD
NINLYRHYNTTDSQTDVYPSGNVQVYELKEGIELFAGLGSARVPPQERYFALD
RMGMENTMDGVNPKLDPERNNELDLGKLTGERYSGLERTFLSYVKDYIVYVRAT
APSSGNMONTKMSYTNIDAYFGYFELSGSVAITDTIFFDGNVAYTRKDDTYPA
KNITDKDIAPLPTARLALRYDTGMVFGVGVLAATQDNVDSDLQEBKTSYGYIIN
LKGGVYRNLRVWAGINNLPDLKYTHLSYLRNPFRTGVKPEPGRYYLSVYTF"
10030..10818
/gene="aq_531"
10030..10818
/gene="aq_531"
/codon_start=1

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.000109	243.00	16256	174
Percent Similarity:	35.90%	Conservative:	115
Best Local Similarity:	21.61%	Mismatches:	301
Query Match:	6.02%	Indels:	216
DB:	1	Gaps:	35

US-09-936-377-2 (1-758) x AE000694 (1-16256)

Qy 29 GlnSerValGlyLeuGluThrValThrValValGlyIlySerArgProGalaThrSer 48

Db 9909 CAGGAAGTGAACCTTGAAGAAATTCAAGTTATCGGAAGAGA----- 9868


```

Db      8086  CTTGAAGGTCGGGTAAATACAGG-----AACCTGAGGCT 8051
Qy      723  llyslala-----AspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSe 741
Db      8050  TGTGCGGGAATTAACAACCTGTTTGAATAACTTACTATACGACCTCTTTATTGAG 7991
Qy      741  xAsp-----ThrProGlnMetGlyArgSerPheThrGlyGlyVa 754
Db      7990  AAATCCCTTCAGAACGGGTGTAAGGTTCCCGAGCCCGAAGGACTTATTATCTTCGCT 7931
Qy      754  lAsnValLysPhe 758
Db      7930  TTCGTACACCTTC 7918

RESULT 40
LOCUS   AE016813
DEFINITION Vibrio vulnificus CMCP6 336182 bp DNA linear BCT 28-FEB-2003
complete sequence.
ACCESSION AE016813
VERSION AE016796
KEYWORDS AE016813.1 GI:27359304
SOURCE Vibrio vulnificus CMCP6
ORGANISM Vibrio vulnificus CMCP6
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 336182)
AUTHORS Rhee,J.H., Kim,S.Y., Chung,S.S., Kim,J.J., Moon,Y.H., Jeong,H. and Choy,H.E.
TITLE Complete genome sequence of Vibrio vulnificus CMCP6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 336182)
AUTHORS Jeong,H., Moon,Y.H. and Kim,J.J.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong, Yuseong-gu, Daejeon 305-811, South Korea
REFERENCE 3 (bases 1 to 336182)
AUTHORS Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2002) Department of Microbiology, Genome Research Center for Enteropathogenic Bacteria, Chonnam National University Medical School, Hak-1-Dong, Dong-Gu, Kwang-Ju 501-746, South Korea
FEATURES
source
1. 336182
/organism="Vibrio vulnificus CMCP6"
/mol_type="genomic DNA"
/strain="CMCP6"
/db_xref="taxon:16895"
/chromosome="II"
288..1034
/locus_tag="VV21363"
288..1034
/locus_tag="VV21363"
/notes="COG0583"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/db_xref="GI:27359305"
/translation="MAERLCLTPGVSHVKKLEVLGSKVKKKKIELTEHGQLL
YDFANDLRKHELNKLQGLVSPVIFISDCLSPTEERKLEELKIGEGSVIHH
GNNVQTVADQLTPVFGDKTLSEFHHNFQILGRDSDYPIDTVMSPLPEKVV
REETHNGRIKIAETCRWIPMGSRIMQYTHMNGRSSLVWCPDQVNSQFASVPLDQV
RFLMKVTHDGLDITAKTLQNVLIESTFTG"
/locus_tag="VV21364"
complement(1101..1934)
/locus_tag="VV21364"
/notes="COG0005"
/codon_start=1
/transl_table=11
/product="Purine nucleoside phosphorylase"

```

```

/protein_id="AAO08250.1"
/db_xref="GI:27359306"
/translation="MTDFVQNALETIYSKPNYQPKVAFILGSGSLGVIADELDDKIV
IFYQELLEPPVSTVVBHSGELVLGKLGIDVIMCKGRGHYYEHQTMKVTNPTFKR
LGEFLLVTAAGSLRPERIDVGLSVFENDHNTWPESTPMGPNDPAGPRFFSLANA
YDKSLRSALTVAEAGEIHLNQGVSFVTFGNFEETAARIMWQIIGGDVGVMSVPEV
ISAAHCGLPVALCAITNMAEGLSDVTLSHAQTLACAKLAADPIKLKAFIKSHFAT
S"
2072..2944
/locus_tag="VV21365"
2072..2944
/locus_tag="VV21365"
/notes="COG0583"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AAO08251.1"
/db_xref="GI:27359307"
/translation="MLRVFDKVAQLGFSSRAEQNLNITKSPLSAQIKLEELVGLTALF
BNTNRVQLTKAGEOLQVECVTLPAQFERSIHRVKQCARERQQLDGLVRFADTRDVAPLLSECLYO
GPAFHALQOLMPHATNLNLELSPEKQVQLLHQJLDGLVRFADTRDVAPLLSECLYO
ECMVVALPSEHPLAQRGLTSLDLKSAFVWLKQENSSTRLIAEYCAKAGYRNIVO
EVVEPNLIAVISAQAVSIVPQSYANLSWPQVRFIPKQRIKQIPAGVYALYHPEAQNSI
KRMIDELKQAIERK"
complement(3014..5149)
/locus_tag="VV21367"
complement(3014..5149)
/locus_tag="VV21367"
/notes="exon:protease Vcc"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AAO08252.1"
/db_xref="GI:27359308"
/translation="MTLOQKRSLILALIFGLSGGSGSGSSPTVQPKDPTPT
TWPDREFTSGQCTATQIADTSSFAQIANADYDCMSWFSSTOIEADALFSEONLA
LVNDALLEVNYQSEERKAKHYIYLAIFKAXYKRDYIARKIAPSTELSONIAT
YSAIFLNTHALEMGDEQKALGSMILIVDSVRQLHARSQVDFALNQFDEQLTENYY
YKAINNIFVAMAGHSQDDFYQLLNDHRYIDTLSTFVSHNQWALGTFDEFLIGNAA
RELARIKTDGDKLEKMTQTLATLLQPPGQSDRVWVGIAEMVNAYPEHAELG
LOQAKERLQRIMTSHDCQGPQAQLTMTLAQAEQVCTLNDEKVDFTVAKTNQO
PPDDNSVDVIVFKTDYTSFSLPNTTNGCQFLERDPSQANNVPRFVAYQN
GWDDPSIINLSEHYVHVLDRFNQYGDPTTMRGHIWVLEGEVYVYKGVHAA
LVLGKXTLTPLEVATSYNDGVNRIYKNGVLGVRLFEKHPKPEVQQLHLSRICEYK
QWVYDEQIGPIYADFESLWDDVTKDIDTVPSPSKKPKPELLALNTFFSVSGAQA
EKLFFVDVAANTRELTITQGGSDADLYACFDKVCYHDIQWINTFHGNSISIQLEKE
SDGYVYKAGRYLISIGREAFVRVTAQA"
5559..5942
/locus_tag="VV21369"
5559..5942
/locus_tag="VV21369"
/notes="COG3094"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein conserved in bacteria"
/protein_id="AAO08253.1"
/db_xref="GI:27359309"
/translation="MMVILLKKAHLGLIVLSFLSALRAYKMGFKASTWLMKEWPLKV
HKVITVMLISALCLTISQYFEADTWLTKELLLVAIVGVAMLAFAKQLNRQRTA
FTSVTCLFAMFYIAKTHVPILS"
complement(5952..7793)
/locus_tag="VV21370"
complement(5952..7793)
/locus_tag="VV21370"
/notes="COG0768; Penicillin-binding protein 2"
/codon_start=1
/transl_table=11
/product="Cell division protein FtsI"
/protein_id="AAO08254.1"
/db_xref="GI:27359310"
/translation="MFARAIATFLILQGLVLIANYLVBNHNFADYQTRSDNRI
KILPIAPVRGLIYDRHGOVLADNILVDELIPEBAENIPDTIKMLKRYISLSDQEA
QFYQRUKQTHFRATIAEDLSEQVARESVHQYQLPGVQISTDPRRYYPYKASLTHT

```


IGVVAHNDHDLTLKEQEKQWYKASRFIKLIEKHYEELLHGOEGYOEVRINSHG
RIIRTKYEPVAGKOLLINIDADLQNYVFTQMAQAGSAIVLSQDNSVLANWSPS
YDNLFSVGSIVSDYRALLNPAHLLRATIGVPPAPVPPKFWALALQVQVITPN
TTRNHWGVRIPDSKTSISRWKMGHGVNVTVAIEESVDSFVAFDLGIDRI
SSWNAFGFRGTIDIDEBTSANMPTRAWKQOTKEEPWYQDTPVIGIGQGYWTATP
LQIAQSTSLVNHGQRTQPHLLYAIKNHGEEMSAIQPVHYDTETMAEVPDIAWISL
NGMLVNHGSKSGSRGAFKAGADYISGKSGTAQVFLAKGETYNSQKLRTLEDHGLE
TGAPOYQHPREVATVVLHGNGGARVCAPLVRKIFDYAIEHKEHPQKG"

gene

/locus_tag="VV21371"

CDS

8309..9151

/locus_tag="VV21371"

/codon_start=1

/transl_table=11

/product="Unknown"

/protein_id="AA08255.1"

/db_xref="GI:27359311"

/translation="MMRLLLILLFSSQSALVLPVVEKVLKQYQAFVNRDPLTI

DDFTGPKARSTVEIVLLOQAFLGVPVLEFHTAPVAERIRRMMLPDYAMAANTI

WDSAVGGDGLWYSDAVIRQGEFAGLYTVSTEDLINEQFPITHDQLSAYSSQWV

SDNWLQALPLSNLNVADWGMVMDVGRHRVDFLLAPFOVSDLSFTTRGFQFPPI

KVKIGLTSRHSRFLSRQFPPESEKLAQALNQIALILRQKGTIERALIECGFINTKVSSW

SKIN"

9324..10076

/locus_tag="VV21372"

9324..10076

/locus_tag="VV21372"

/note="COG3201"

/codon_start=1

/transl_table=11

/product="Nicotinamide mononucleotide transporter"

/protein_id="AA08256.1"

/db_xref="GI:27359312"

/translation="MELFSLFDINNTLVNIPIDGVAMSWIEAIGTIFGLLCIFPASQ

KXTNYLGINVITLFEVIFQIQLYGLLLQLFFCAVYGYWYATRNPAEGDLELV

RWLSQKQWIPAGLCVCAIALTVYIDPFFALANIAVDVMTNVEFGAGLAEPVLEPDAF

PFDSTMTVLSIAQILMTKRYVENVWLVVWINISVGIYAMQGVYVMSIQYAILMFI

AANGTREWAFKRNQASTQKSEKSADEAVA"

10073..10801

/locus_tag="VV21373"

10073..10801

/locus_tag="VV21373"

/note="COG2820"

/codon_start=1

/transl_table=11

/product="Uridine phosphorylase"

/protein_id="AA08257.1"

/db_xref="GI:27359313"

/translation="MSKQPHLCVDETQVAERVIVCGEPDRANRIFAALNAELVAENR

EYRLFNGEFOGKAITVCSGTGIGAPSMIIAVEELKQCGAKYIIRVGSAGAMQPNIALGE

LIVAEAGVRDGGSKAVIDAAYPAYAFSLKGLDGLDYLOTQAVPYHMGVVRSHDSFT

DDEETLCQYNNKGIIGADMETSALFTVGRGLRLHLVASVLNNVRYQQDVKEGVSQYV

NDDDMWAGEKRASLALALSCQ"

complement(10989..11585)

/locus_tag="VV21374"

complement(10989..11585)

/locus_tag="VV21374"

/note="Other COG candidate exists; COG3326"

/codon_start=1

/transl_table=11

/product="Predicted membrane protein"

/protein_id="AA08258.1"

/db_xref="GI:27359314"

/translation="MAIKQIIEWNDERKGVFTSAIGELKVFFHSSVTNRGVRPKI

KDNVFNVAEDKGRFNAENVVLGVHGFPPFTVLFPSFFVAATASVFLFDGPKILIP

LYVLSMFTYLMFAKDKQAQADQGRWRTPESTLHLLSLGGWPGALLAQFLLRHKSQKQ

PFKFLMLTITLITAGFLWLFETSGRHLVQVVFCAEFL"

11843..12523

/locus_tag="VV21376"

11843..12523

/locus_tag="VV21376"

/locus_tag="VV21376"

/locus_tag="VV21376"

/locus_tag="VV21376"

/locus_tag="VV21376"

Alignment Scores:

Pred. No.:	0.00721	Length:	336182
Score:	240.50	Matches:	173
Percent Similarity:	35.87%	Conservative:	124
Best Local Similarity:	20.89%	Mismatches:	344
Query Match:	5.96%	Indels:	187
DB:	1	Gaps:	37
US-09-936-377-2 (1-758) x AB016813 (1-336182)			
QY	1	MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuAenThrProLeu	20
DB	188949	ATGAACCTAAACGCGCTGCTCAGCGCGGTACTGTCGTACTCGCTGCCAATCTCGTGCAC	189008
QY	21	LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly	40
DB	189009	GCCGAATATGCCACGCGCGTATGGAAGAGGTC-----GTCTGTGACGGCTAATCAA	189059
QY	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
DB	189060	ATCGAGCAACCG-----CTCTCAGAACTGGCGGCTCTGTGCGGTG	189101
QY	61	IleSerGlyAspThrLeuArgGlnLysAla-----ValAsnLeuGlyAspAlaLeuAspGly	79
DB	189102	ATCAGTGTAAAGAACTAGAGAAAAAGGCCAACACAGAGCTGTACGATGCCCTCGCGCAAT	189161
QY	80	ValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaPro-----Val	96
DB	189162	GAGCCCGCGCTCAGTGTACAG-----GGAGCGCGCGGTGCGGCACAAACATCGCC	189212
QY	97	IleArgGlyGlnThrGlyArgGlyIleLysValLeuAsnHisHisGlyGluThrGlyAsp	116
DB	189213	ATTCGTGTATGACCGGCAACCAATATATGATCGTTCGCGATGCGGTGCGAAGCGCGAT	189272
QY	117	MetAlaAspPheSerProAspHis-----AlaIleMetValAspThr	130
DB	189273	-----GGTITGTGTGCCAATGACATCAATGACAAAGTGGGGCGTAATTCCTTTGATTG	189326
QY	131	AlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGly	150
DB	189327	GTCACAGCTCAAAAGTATTGAAGTAGTAAAGAGCCAGTTCATCATTCACGGCTCAGCG	189386
QY	151	AsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn	170
DB	189387	GCACTGGGTGGCACCCTGTC-----GTGATTGAATCGATGACCGCGGTACTACTTAA	189443
QY	171	-----GlyValSerGlyGluLeu-----	176
DB	189444	GACGATTTCTATGCTGTGATGTGGCCACACGATACAGGCAATCAGTGATAAAACCACTG	189503
QY	177	-----GlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsn	194
DB	189504	ATGAGTAAATTTGGCCTTTCGCTCAGGCAACACCGCCTGTTGCTGACG-----GT	189557
QY	195	IleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAsp	214
DB	189558	TATTCGCAAGGAAGAAACCGCAACTCAATCAAGATACTACCAACGTCAGTGTAGAA	189617
QY	215	TyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAla	234
DB	189618	GCGGCGAGCGCGCAGCTACAGCTCAACCATTTCTGTTGTTGATGCG-----	189662
QY	235	AsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSer	254
DB	189663	-----CTGATGTTAAGTGTTCGCGCAGATGCTTATCGTGAAGGGCAAAAT	189707
QY	255	AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla	274
DB	189708	CGTCTGTAAGGAGCTCAGGTATA-----CATCGC	189737
QY	275	AspIleIleTrpGlnLysSerLeuLeuAsnLysArgTyrLeuGlnLeuTyrProHisLeu	294
DB	189738	GATGCAAAATGGATCAAAAC-----CAGTATTAC-----	189767

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 16:19:56 ; Search time 366 Seconds
(without alignments)
5590.639 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTLKPIVLISILLINPL.....FLSDTPQMGRSFGVGVNKF 758

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO spool/US09936377/runat 14112003 104449 22098/app query.fasta_1.903
-DB=N_GenSeq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOEFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936377 @CGN 1.1.0 @runat 14112003 104449 22098 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_GenSeq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4036	100.0	2277	21	DNA encoding a Nei
2	3828	97.3	2295	21	Neisseria meningit
3	3828	97.3	2295	24	Nucleotide sequenc
4	3827	97.3	2277	24	Neisseria meningit
5	3827	97.3	349980	21	Neisseria meningit
6	3827	97.2	2277	21	Neisseria meningit
7	3822	97.2	2277	21	Neisseria meningit
8	3822	97.2	1437668	21	N. meningitidis B
9	3820	97.1	2277	21	Neisseria meningit
10	3812	94.4	2208	21	Neisseria gonorrhoe
11	3812	94.4	22863	21	Neisseria meningit
12	3706	91.8	2184	25	N. meningitidis pa
13	1808	44.8	1020	21	N. gonorrhoeae nuc
14	1791	44.4	1020	21	Neisseria meningit
15	1763	43.7	1020	21	Neisseria meningit
16	1585	39.3	2457	22	Neisseria gonorrhoe
17	1585	39.3	2846	24	Moraxella catarrha
18	1585	39.3	89047	22	Moraxella catarrha
19	1205.5	29.9	2942	24	Genomic fragment #
20	1205.5	29.9	269223	22	Genomic fragment #
21	1202	29.8	2742	22	Moraxella catarrha
22	803	19.9	522	21	Neisseria gonorrhoe
23	651.5	16.1	2127	24	Neisseria meningit
24	271.5	6.7	1173	22	M. capsulatus gene
25	230	5.7	1830121	17	Pseudomonas sp hea
26	229.5	5.7	2145	24	Haemophilus influe
27	229.5	5.7	26778	21	Neisseria meningit
28	229.5	5.7	349980	21	N. meningitidis pa
29	229.5	5.7	837096	21	Neisseria meningit
30	227.5	5.6	2307	24	N. meningitidis pa
31	226.5	5.6	2586	22	Nucleotide sequenc
32	223	5.5	2139	25	Pseudomonas sp hea
33	223	5.5	2368	17	N. gonorrhoeae nuc
34	223	5.5	2381	22	Neisseria gonorrhoe
35	220.5	5.5	32160	24	Neisseria gonorrhoe
36	214	5.3	4118	17	B. coli CFT073 gen
37	212.5	5.3	2112	21	NTHI hxC + hxCB g
38	212.5	5.3	2112	22	Neisseria meningit
39	212.5	5.3	8367	21	Neisseria meningit
40	212.5	5.3	349980	21	N. meningitidis pa
41	212.5	5.3	349980	21	Neisseria meningit
42	212.5	5.3	1437668	21	Neisseria meningit
43	208.5	5.2	2112	21	N. meningitidis B
44	208.5	5.2	2112	22	Neisseria meningit
45	207	5.1	2600	17	Neisseria meningit

ALIGNMENTS

RESULT 1

AAA75744
ID AAA75744 standard; DNA; 2277 BP.
XX AAA75744;
AC
XX
22-JAN-2001 (first entry)

DNA encoding a Neisseria meningitidis BASB082 polypeptide.

BASB082; BASB083; BASB091; BASB092; BASB0101; infection; vaccine;
gene therapy; upper respiratory tract infection; bacteremia; meningitis;
invasive bacterial disease; ss.

Neisseria meningitidis.

XX Key Location/Qualifiers

XX Key Location/Qualifiers

```

FT 1..2277
FT /*tag= a
FT /product= "BASB082"
XX
PN WO200055327-A2.
PD 21-SEP-2000.
XX
XX 07-MAR-2000; 2000WO-EF01955.
XX
XX 12-MAR-1999; 99GB-0005815.
XX 21-APR-1999; 99GB-0002094.
XX 23-APR-1999; 99GB-0009503.
XX 28-APR-1999; 99GB-0009787.
XX 07-MAY-1999; 99GB-0010710.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Defrenne C, Delmelle C, Ruelle J;
XX
XX WPI; 2000-602119/57.
XX P-PSDB; AAB18719.
XX
XX Novel polypeptides designated BASB 082, 083, 091, 092, and 101 derived
XX from meningococcus bacterium useful for producing vaccines against
XX infections and in diagnostic assays -
XX
XX Claim 11; Page 98-99; 108pp; English.
XX
XX The present sequence encodes a BASB082 polypeptide. The specification
XX describes BASB082, BASB083, BASB091, BASB092, and BASB101 polypeptides.
XX The polynucleotides and polypeptides are useful as diagnostic reagents
XX and for diagnosing N. meningitidis infection. The polynucleotides may
XX be used as hybridisation probe for RNA, cDNA and genomic DNA to isolate
XX full-length cDNAs and genomic clones encoding BASB082, BASB083, BASB091,
XX BASB092 or BASB101 polypeptides and to isolate cDNA and genomic clones
XX of other genes that have a high identity particularly high sequence
XX identity to BASB082, BASB083, BASB091, BASB092 or BASB101 genes. The
XX vaccine compositions are useful for inducing an immunological response
XX in humans. The polynucleotides encoding BASB082, BASB083, BASB091,
XX BASB092 or BASB101 polypeptides are useful in gene therapy to induce
XX an immunological response. The polypeptides are useful for treating
XX upper respiratory tract infection, invasive bacterial diseases, such as
XX bacteremia and meningitis.
XX
XX SQ Sequence 2277 BP; 596 A; 713 C; 553 G; 415 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 0 Length: 2277
Score: 4036.00 Matches: 758
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

```

US-09-936-377-2 (1-758) x AAA75744 (1-2277)

```

QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
DB 1 ATGGCACAACCTACATCAACCCATGTTTATCATCTTTTATCAACACACCCCTC 60
QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
DB 61 CTGCCCCAACGGCATGAACCTGAGCAATCGTGGGCTTGGAACCGGTTACCGCTCGCGC 120
QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
DB 121 AAAAGCCGTTCGGCGGCCACATCAGGGCTCTGCACACTTCGACCCCTCCGACAAAATC 180
QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
DB 181 ATCAGCGCGACACCTTGCACAAAAGCCGTCACCTTGGGGGATGCTTTGACGGCGTG 240

```

```

QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
DB 241 CCGGGCATTACGCTCGCAATACGGCGCGCGCGCTCCGCTCCGTTATTTCGGGTCAA 300
QY 101 ThrGlyArgArgIleIleValLeuAsnHisIleGlyGluThrGlyAspMetAlaAspPhe 120
DB 301 ACAGGCAGACCGGATTAAAGTATTGAACCATCACGGCGAAACAGGCGATATGGCGATT 360
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValIleGluLeuArg 140
DB 361 TCGCCCGATCACGCCATTATGATAGTACCGCTTGTCTGCAACAGTCCGAAATCTTCGC 420
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
DB 421 GGGCCGGTTACGCTCTTGTACAGCTCGGCGCAATGTGGCGGGGCTGTGTCGATGTTG 480
QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
DB 481 GSCAAAATCCCCGAAAAAATGCTGAAAAGCGCGTATCGGGCGAACTCGGATTGGCTTTG 540
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
DB 541 AGCAGCGGCAATCTGAAAAAATCTCAGCTCCGGCGGCATCAATATCGGTTTGGGCAAAAAC 600
QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
DB 601 TTTGTATTGCACACGGAAGGGCTGTACCGCAATCGGGGATTACCGCGTACCGCGTTAC 660
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
DB 661 CGCAATCTGAAACGCGCTGCCGACAGCCACCGCATTCGCAACAGGCGGCATCGGCT 720
QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260
DB 721 GTCTTGGTGGCGAAAAAGGTTTATCGCGCTAGTACAGCAACGTCGCGCAACATAT 780
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
DB 781 GGTCTGCTGCCGACAGCCAGCAATACGATGATGTCACGCGCGACATCATCTTGGCAAAAG 840
QY 281 SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVal 300
DB 841 AGCTTGAATTAACAAACGCTATTATACGCTTTATCGCACCTGTTGACCGAAGAACGCTC 900
QY 301 AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
DB 901 GATTACGACAAATCCGGGCTTGAGCTGGCGCTTCCACGACGACGATGTCACACGCCCAT 960
QY 321 AlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
DB 961 GCCCACAACGGCAACCTTGGATAGACTTCGCGCAACAAACGCTACGACTCCGCGCGGAA 1020
QY 341 TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
DB 1021 TGAAGCAGCCATTCGCCGCTTTTGAAGCCCTGCGCGTACACCTGAACCGCAACGACTAC 1080
QY 361 HisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsn 380
DB 1081 CACACGACGAAAAACGCGGATGCGTGCAGAAACTTTTTTAAACACCAACCGCAAAAC 1140
QY 381 AlaArgIleLeuLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGln 400
DB 1141 GCCCGCATCGAGTTGGCGCACCAACCCATAGGCCCTCTGAAAGGCGAGCTGGGCGGTGCA 1200
QY 401 TyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet 420
DB 1201 TATTGGGCAAAAAATCCAGTGCTTTATCTGCCACATCCGAAAGCGGTCAAAACCGCATG 1260
QY 421 LeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValIleGluAlaAsnTrpAsp 440
DB 1261 CTGCTTGACAAATAAGTGCAACATTACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGAC 1320
QY 441 AsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAsp 460

```

Db 1321 AACTTACGCTTGAAGCGCGTACGCGTGAAGAAACAAAAGCCTCCATCGCTACGAC 1380
 Qy 461 LysAlaLeuLeuLeuAspArgGluAenTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
 Db 1381 AAGCATTTGATTCGGGAAACTACTACAGCAGCCCTGCGGACCTCGCGCGGAC 1440
 Qy 481 ArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLys 500
 Db 1441 CGCCAAACCGCCGCTCGTTGCGACCTTCGGGCAACTGGTATTTACGGCGGACAA 1500
 Qy 501 LeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 Db 1501 CTCAGCCTGACCGCTCCCATCAGGAAGCGCTGCGCTACCGCAAGAGCTGTACGCAC 1560
 Qy 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
 Db 1561 GGCAAAACAGGTGGCCCAACACTTTTGAAGTCGGCAACAACTGGAACAAAGAGCT 1620
 Qy 541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla 560
 Db 1621 TCCAAACAATCGAACTCGGCTTGGCTACGAAGCGGACCGCTGGCAATACAATCTGGCA 1680
 Qy 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
 Db 1681 CTCCTACCGCAACCGCTTCGGCACTCATTTACGCCCAACCTTAACGACGACGCGGC 1740
 Qy 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAla 600
 Db 1741 CCCAAATCCATCGAAGACGACAGCGAAATGAAGCTGCTGCGCTACAAACCAATCCGCTGG 1800
 Qy 601 AspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGly 620
 Db 1801 GACTTCTACCGCGCGAAGCGAAATCTTACTTCAACCGGACCGCGCTACCGCATCGGC 1860
 Qy 621 ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGlu 640
 Db 1861 GTTTCGCGGACTATGATACGAGGCGCTGAAAAACCTGCGCTCCCTACCGCGCAGGAA 1920
 Qy 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
 Db 1921 GATGCTTACGGCAACCGCTCTTTCATCGCGGAGGACGACCAAAACCGCCCTCGCTCCG 1980
 Qy 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
 Db 1981 GCTGCGGCTCGCTTCCACTGAAAGCTCGCTGACCGCGCATCGATGCAATTTG 2040
 Qy 681 AspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
 Db 2041 GACTACTACCGCGCTTTTGGCCCAAAACAACTCGCCGCTACGAAACGCGCAGCCCGGA 2100
 Qy 701 HisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsn 720
 Db 2101 CACCATATGCTCACTTCGGCGCAAACTACCGCGCAATACCGCTATGCGAGTGAAT 2160
 Qy 721 TrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeu 740
 Db 2161 TGGTACGTCGCAAGCGCAACCTGCTCAACCAATTCGCTTTACGCCCAACGACGCTTCTC 2220
 Qy 741 SerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2221 TCTGATACGCCCAAAATGGCGCGGACGCTTACCGGTGGCTTAACGTGAAGTTT 2274
 RESULT 2
 ID AAZ53319
 AC AAZ53319; standard; DNA; 2295 BP.
 AC AAZ53319;
 DT 21-MAR-2000 (first entry)
 XX Neisseria meningitidis ORF 147 partial DNA sequence SEQ ID NO:587.
 DE
 XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 XX Neisseria meningitidis.
 XX WO957280-A2.
 XX 11-NOV-1999.
 XX 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR P-PSDB; AAY74557.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PT Claim 7; Page 414; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX SQ Sequence 2295 BP; 606 A; 724 C; 553 G; 412 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 2295
 Score: 3928.00 Matches: 744
 Percent Similarity: 98.68% Conservative: 5
 Best Local Similarity: 98.02% Mismatches: 9
 Query Match: 97.32% Indels: 2
 DB: 21 Gaps: 0
 US-09-936-377-2 (1-758) x AAZ53319 (1-2295)
 QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuLeuLeu 20
 Db 19 ATGGCACAACCTACACTCAACCCCATTTGTTTATCAATTTCTTTTAAACAACACACCCCTC 78
 QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 Db 79 CTCTCCCAAGCGCATGGAACTGAGCAATCAGTGGGCTTGGAAACGTCAGCGTCGTGGC 138
 QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 139 AAAAGCGCTCGCGCGCCACTTCGGGGCTGTCGACACTTCTACCGCTTCGCAAAATC 198

QY 11leSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 199 ATCAGCGCGACACCTTGGCACAACAAAGCCGTCACCTTGGGTGATGCTTTAGACGGCGTA 258
QY 81ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db 259 CCGGGCATTCATGCTCGCAATACCGCGCGCGCATCCGCTCCGCTTATTCGGGGTCAA 318
QY 101ThrGlyArgArgIleLysValLeuAsnHisIleSgIyGluThrGlyAspMetAlaAspPhe 120
Db 319 ACAGGCGACAGGATTAAGTGTGAACCATCAGCGCGAAACGGGCGACATGCGCGCACTTC 378
QY 121SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140
Db 379 TCTCCAGACCATGCAATCATGTTGACAGCTCGGCAATGTGGCGGGCTGGTGCATGTTGCCGAT 438
QY 141GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 439 GGTCCGGTTTACGCTCTGTGTACAGCTCGGCAATGTGGCGGGCTGGTGCATGTTGCCGAT 498
QY 161GlyIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
Db 499 GGCAGAAATCCCGAAATAATGCTGAAACGCGGTATCGGGCAACTCGGAATGCGGTG 558
QY 181SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db 559 AGCAGCGGCAATCTGGAAATACTCAGCTCGGCGGCATCAATATCGGTTTGGGCAAAAC 618
QY 201PheValLeuHisThrCluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 619 TTTGTATTGCACAGGAAGGGCTGTACCGCAATCGGGGATTACCGCGTACCGGGTTAC 678
QY 221ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 679 CGCAATCTGAAACCGCTGCGCGACAG-CCACGCGGATTCGCAACCGGCGAGCATCGGCT 737
QY 241ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
Db 738 GTCTTGGTGGGAAAGGGCTTTATCGCGCGACATACAGCGACCGCTGCGGACCAATA 797
QY 260rGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
Db 798 TGGTCTGCTGCCACAGCCAGCAATACGATGATTGCCAGCGCGACATCATCTGSCAAA 857
QY 280sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysPva 300
Db 858 GAGTTTGTATTAACAAACGCTATTTCAGGCTTTATCGGACCTGTTGACCGAAGACAT 917
QY 300lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
Db 918 CGATTACGACAAATCCGGGCTTGAGCTGGGCTTTTCACGACGACGATGATGCACACGCCA 977
QY 320sAlaHisAsnGlyLysProTyrIleAspLeuArgHisLysArgTyrGluLeuArgAlaGl 340
Db 978 TGCCCAACACGGCAACCTTGGATAGACTCGGCAACAAACGCTACGAACTCCGCGCGGA 1037
QY 340uTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
Db 1038 ATGGAAGCAACCGTTCCTCCGGTTTGAAGCCCTGGCGGTACACCTGAAACCGCAACGACTA 1097
QY 360rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
Db 1098 CCGCCACGACGAAAGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1157
QY 380nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrrGlyValGl 400
Db 1158 CGCCCGTATCGAGTTGCGCACCAACCCATAGGCGCTGTAAGAGGCGAGCTGGGGGTGCA 1217
QY 400nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
Db 1218 ATATTTGGGCAAAATCCAGTCTTTATCTGTCACATCCGAGCGGTCAACCAACCGCAT 1277

QY 420tleuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAs 440
Db 1278 GCTGCTGCATATAAGTGCAACATTACAGCTTTTTCGGTGTAGAACAGGCAACTGGGA 1337
QY 440PhePheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db 1338 CAATTCACGCTTGAAGCGGGGTACGGTGGAAAAACAAAAAGCCTCCATCCGTACGA 1397
QY 460pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
Db 1398 CAAAGCATTTGATTCGGGAAAACTACTCAACCATCCCTGCGCGACCTCGCGCGCA 1457
QY 480sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLys 500
Db 1458 CGGCCAAACCGCGCTCATTCGCACTTTCGGCACTTGGTATTTTCAACGCCAACACAA 1517
QY 500sLeuSerLeuThrAlaSerHisGlnArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 1518 ACTCAGCCTGACCGCTCCCATCAGAACCGCTGCGCTCAACGCAAGAGCTGTACGCACA 1577
QY 520sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
Db 1578 CGGCAACACGTCGCCCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG 1637
QY 540gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrrGlnTyrAsnLeuAl 560
Db 1638 TTCACAAATATCGAACTCGCGCTGCTACGAGCGACCGCTGGCAATCAATCTCGGC 1697
QY 560aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
Db 1698 ACTCTACCGCAACCGCTTCGGCACTCATTTTACGCCCAACCTTAAACGCGACGCGCG 1757
QY 580YProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db 1758 CCCCACAAATCCATCGAAGACGACAGCAATGAAGCTCGTGGCTTACACCAATCCGGTGC 1817
QY 600aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl 620
Db 1818 GGAATCTTACCGCGCGGAAGGCAATCTTACTTCAACGACACCGCGCTACCGCATCGG 1877
QY 620YValSerGlyAspTyrValArgIleValGluLysAsnLeuProSerLeuProGlyArgGl 640
Db 1878 CGTTTCCGGCGACTATGTACGAGGCGCTCTGAAAAACCTGCTTCCCTACCGCGCAGGA 1937
QY 640uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValer 660
Db 1938 AGACGCTTACGGCAACCGGCTTCAATTCGCCAAGCCGACCAACGCCCCCTCGCGTTC 1997
QY 660oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
Db 1998 GGCTGCGCGCTCGGCGTCCACTCGAAAGCCTCGTACCGACCGCATCGATGCCAATT 2057
QY 680uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
Db 2058 GGACTACTACCGCGTTCGCCCAACAACTTCGCCCGCTACGAAACGCGCACCGCCCCGG 2117
QY 700YHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyLysTrpAs 720
Db 2118 ACACCATATGCTCAACCTCGGCGCAACTACCGCGCATATCGCGTATGGCGAGTGGNA 2177
QY 720nTrrTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
Db 2178 TTGGTACGTCAAGCGCGCAACCTGTCTCAACCAATCCGTTTACGCCCAACGACGAGCTTCT 2237
QY 740uSerAspThrProGlnMetGlyAspSerPheThrGlyGlyValAsnValLysPhe 758
Db 2238 CTCTGATACGCCCAATTCGGCGCGAGCTTACCGCGCGGTGAACGTGAAGTTT 2292

RESULT 3

ABQ78299

ID ABQ78299 standard; DNA; 2295 BP.

XX

AC ABQ78299;

Db 1278 GCTGCTTGAACAATAAAGTGCACATACAGCTTTTCGGTGTAGACAGGCAAACTGGGA 1337
 Qy 440 pAsnPhrThrLeuGluGlyClyValArgValGluGlyGlnLysAlaSerIleArgTyrAs 460
 Db 1338 CAACCTTCACCGCTTGAAGCGCGGTAGCGGTGGAAAAACAAGAAGCTTCATCCGTACGA 1397
 Qy 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
 Db 1398 CAAAGCATTCATTGATCGGGAATACTACAAACCATCCCTCCCGACCTCGGCGCGCA 1457
 Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLys 500
 Db 1458 CCGCAAAACCGCGCTCATTCGCACCTTCGGGCAACTGGTATTCACGCCCAACACAA 1517
 Qy 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 Db 1518 ACTCAGCCTCAGCGCTCCCATCAGGAACCGCTCGCGTCAACGCAAGAGCTGTACGCACA 1577
 Qy 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAspLysGluAr 540
 Db 1578 CGGCAAAACAGCTGCCACCAACCTTTGAAGTCGGCAACAACACCTCAACAAAGAGCG 1637
 Qy 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
 Db 1638 TTCCACAATATCATCGCTCGCTGGGCTACGAAGCGACCGCTGGCAATACAACTCTGGC 1697
 Qy 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
 Db 1698 ACTTACCGCAACCGCTTCGGCAACTACATTTAGCGCCAAACCTTAAACGACGAGCGG 1757
 Qy 580 yProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 Db 1758 CCCCAAATCCATCAGACGACGAGCAAAATGAAGCTCGTGGCTACAACCAATCCGGTGC 1817
 Qy 600 aAspPheTyrClyValGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl 620
 Db 1818 GCACTTCACGGCGCGAAGCGAAATCTACTTCAAACCGACACCGCGCTACCGCATCGG 1877
 Qy 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGl 640
 Db 1878 CGTTTCGGCGCACTATGTACGAGCGCGCTCTGAAAAACCTGCTTCCCTACCGGACAGGA 1937
 Qy 640 uAspAlaTyrClyValAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 Db 1938 AGACGCTTACGGCAACCGCCCACTCATGTGCCAAGCGCAACCAACGCGCTTCGCTTCC 1997
 Qy 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
 Db 1998 GGCTGCGCGCTCGCGCTCCACCTGAAAGCTCGCTGACCGACCGCATCGATGCCAATTT 2057
 Qy 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
 Db 2058 GGACTTACTACCGCGTGTTCGCCCAAAACAACTCGCCGCTACGAAACGCGACGCGCGG 2117
 Qy 700 yHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluThrPAs 720
 Db 2118 ACACCATATGCTCAACTCGCGGCAAACTACCGCGCAATACCGGCTATGGGAGTGGAA 2177
 Qy 720 nTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 2178 TTGGTACGTCAAGCGCAACCTCTCAACCAATCCGTTTACGCCCAACGAGCTTCCT 2237
 Qy 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2238 CTCGTATACGCGCAAAATGGGCGCGAGCTTACCGGCGCGGTGAACGTGAAGTTT 2292

RESULT 4

AAS97296

ID AAS97296 standard; DNA; 2277 BP.

XX

AC AAS97296;

XX

12-MAR-2002 (first entry)
 Neisseria meningitidis virulence gene #101.
 Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
 infection; Gram-negative bacteria; antimicrobial; ds.
 Neisseria meningitidis.
 WO200185772-A2.
 15-NOV-2001.
 08-MAY-2001; 2001WO-GB02003.
 08-MAY-2000; 2000GB-0011108.
 (MICR-) MICROSCIENCE LTD.
 Tang C;
 WPI; 2002-066593/09.
 P-PSDB; AAU73011.
 New peptide encoded by operon including virulence genes of Neisseria meningitidis, useful as vaccine component for treating or preventing meningitis and for identifying antimicrobial drug -
 Claim 1; Page 408-412; 423pp; English.
 The invention relates to a peptide (I) encoded by an operon (II) of Neisseria meningitidis including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in a Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by Neisseria or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAS97196-AAS97305 represent N. meningitidis virulence genes and related PCR primers of the invention.
 Sequence 2277 BP; 600 A; 712 C; 546 G; 419 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 2277
 Score: 3927.00 Matches: 743
 Percent Similarity: 98.81% Conservative: 7
 Best Local Similarity: 97.89% Mismatches: 8
 Query Match: 97.30% Indels: 2
 DB: 24 Gaps: 0

US-09-936-377-2 (1-758) x AAS97296 (1-2277)

Qy 1 MetaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
 Db 1 ATGGCAACAACACTACACACCAACCACTTTTATCAATCTTTTATCAACACACCCCTC 60
 Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 Db 61 CTCGCCCAAGCGCATGAATGAGCAATCGTGGATTTGAAACGGTCAGCGTCGCGC 120
 Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 121 AAAAGCGCTCCGCGCGCACGTCCGGGCTGTTGCACACTTCGACCGCTCCGACAAATC 180
 Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 181 ATCTCCCGCGATACCTTCGCGCAAAAGCGCTCAACTTCGGCGACGCTTTTACACGCGTA 240
 Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyValAlaSerAlaProValIleArgGlyGln 100
 Db 241 CCGGGCATCCACGCTTCGCAATACGGCGCGGCGGCTCTGCTCCCTCATTCGGGTCAA 300

QY 101 ThrGlyArgArgGlyLeuValLeuAsnHis sH sGlyGluThrGlyAspMetAlaAspPhe 120
 Db 301 ACAGCGCGCGGATTAAGTGTGAACCATCACGCGGAAACAGGCGATATGCGGATTTT 360
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 361 TCGCCCGATCAGCCCATATGTAGATACCGCTTGTGCGAACAGGTGCGAATCCTCGGC 420
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyLeuValAlaGlyLeuValAspValAlaAsp 160
 Db 421 CGGCGCGTGTACGCTCTGTACAGCTCGGCGCAATGCGCGGCGCTGGTGTGATGTTGCCGAT 480
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 481 GGCANAATCCCGAANAATGCTGAAACGCGATTCGCGCGAATTCGGATTGCGGTTG 540
 QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 541 AGCAGCGGCAATCTGGAANAATCTACGTCGCGCGCATCAATATCGGTTTGGGCAAAAAC 600
 QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 601 TTGTATTGCAACAGGAGGCTGTACCGCAATCGGGGATTACGCGGTACCGGTAC 660
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 661 CGCAATCTGAAACGCGCTGCGGACAG-CCACGCGCATTCGCAACGCGGACGATCGGCGT 719
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
 Db 720 GTCTTGGTGGCAAAAAGTTTATCGCGTAGCTAGCTACGCGACCGTCCGGACCAATA 779
 QY 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
 Db 780 TGGTCTGCTGCCACAGCGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLeuAspVal 300
 Db 840 GAGCTTGATTAACAACGCTATTACAGCTTTATCGCACCTGTGACCGAGAGAGACAT 899
 QY 300 AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
 Db 900 CGATTAACGCAATCCGCGGCTTGAGCTGCGGCTTCCACGACGACGATATGACACACGACA 959
 QY 320 sAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGly 340
 Db 960 CACCCACAGCGGACGCTGTATGATGACCTGCGCAACAAACGCTACGAACTCCGTCGCGCA 1019
 QY 340 uTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
 Db 1020 ATGGAAGCAACCGTTCCCGGTTTGAAGCCCTTGGCGGTACACCTGAAACCGCAACGACTA 1079
 QY 360 rHisHisAspGlyAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 Db 1080 CCGCCACGACGAAAGAGCGCATGCGATCGAANAATTTTTTAACACCAACGCAAAA 1139
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrrPglyValGly 400
 Db 1140 CGCCCGCATCGAGTTGCGGCACCAACCCCATAGTGTCTGAAAGGCGAGTGGGCGTGC 1199
 QY 400 nTrrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet 420
 Db 1200 ATATTACAAANAATCCAGTCTTTATCTGCCATATCCGAAGCGGTAAACACACCGAT 1259
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrrPAs 440
 Db 1260 GCTGCTTGACAAACAAAGTCAACATACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGA 1319
 QY 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 Db 1320 CAACTTCACGCTTGAGAGGAGCGTACGCGTGGAAAAACAAAGGCTCCATTCAGTACGA 1379

QY 460 pLysAlaLeuIleAspArgGlyAsnTrrTrrLysGlnProLeuProAspLeuGlyAlaHis 480
 Db 1380 CAAAGCATTTGATGTGCGAANAATCTACTCAACACCCCTGCGGACCTCGCGGCGCA 1439
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrrPheThrProGlnHisLys 500
 Db 1440 CGCCCAAAACCGCGCTCATTCGCACTTTTCGGCAACTGTAATTTTCACGCGCACACAA 1499
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTrrAlaHis 520
 Db 1500 ACTAGCCTTGACGCGCTCCCATCAGAACCGCTGCGGTCAACGCAAGAGCTGTACGACA 1559
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
 Db 1560 CGGCAAAACAGCTGCGCACCAACCTTTGAAGTCGCGCAACAAACCTCAACAAAGAGCG 1619
 QY 540 sSerAsnAsnIleGluLeuAlaLeuGlyTrrGluLysAspArgTrrPheGlnTrrAsnLeuAl 560
 Db 1620 TTCCCAACAATATCGAATCGCGCTGCGGTACGAGCGGACCGCTGGCAATCAATCTCGC 1679
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTrrIleTrrAlaGlnThrLeuAsnAspGlyArgGly 580
 Db 1680 ACTCTACGCAACCGCTTCGGTAACTACATTTACGCCCAAACTTTAAACGCGGACGCG 1739
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTrrAsnGlnSerGlyAl 600
 Db 1740 CCCCAAATCCATCGAAGACGACGAGGAAATGAAGCTCGTGGCTACAAACCAATCCGCGCG 1799
 QY 600 aAspPheTyrGlyAlaGluGlyLysIleTrrPheLysProThrProArgTrrArgIleGly 620
 Db 1800 CGACTTCTACGCGCGGAGGCGAATCTACTTCAACCGGACACCGCTACCGCATCGG 1859
 QY 620 yValSerGlyAspTrrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGly 640
 Db 1860 CGTTTCCGCGCATGTGTACGAGGCGCTGTGAAAAACCTGCTTCCCTTACCGGCGAGGA 1919
 QY 640 uAspAlaTrrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPrr 660
 Db 1920 AGATGCTACGCAACCGCTTCTTCATCGCACAGACGACCAAAATGCCCCCGGTCTCC 1979
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
 Db 1980 GGTGCGCGCTCGGCTTCCACTGAAAGCTTCGCTGACCGACCGTATCGATCGCAATTT 2039
 QY 680 uAspTrrTrrArgValPheAlaGlnAsnLysLeuAlaArgTrrGluThrArgTrrProGly 700
 Db 2040 GSACTACTACGCGGTGTTGCGCCAAAACAACTCGCCCGCTACGAAACGCGCACGCGCGG 2099
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTrrArgArgAsnThrArgTrrGlyGluTrrPAs 720
 Db 2100 ACACCATATGCTCAACCTCGCGCGCAAACTACCGCGCAATACGCGCTATGCGAGTGGAA 2159
 QY 720 nTrrTrrValLysAlaAspAsnLeuAsnGlnSerValTrrAlaHisSerSerPheIle 740
 Db 2160 TTGGTACGTCAAAGCGCAACCTGCTCAACCAATCCGTTTACGCCCAACAGCGCTTTCT 2219
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2220 CTCTGATACGCGCAATGCGCGGAGCTTTTACCGCGCGGTGAACGTGAAGTTT 2274

RESULT 5

AAF21609/c

ID AAF21609 standard; DNA; 349980 BP.

XX AAF21609;

AC AAF21609;

DT 13-MAR-2001 (first entry)

DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;

KW ds.

XX OS Neisseria meningitidis.
 XX PN WO200066791-A1.
 XX PD 09-NOV-2000.
 XX PF 08-MAR-2000; 2000WO-US05928.
 XX PR 30-APR-1999; 99US-0132068.
 XX PR 08-OCT-1999; 99WO-US23573.
 XX PR 28-FEB-2000; 2000GB-0004695.
 XX PA (CHIR) CHIRON CORP.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
 XX PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;
 XX PI Frazer CM, Grandi G;
 XX PX WPI; 2000-647603/62.
 XX DR
 XX XX Neisseria meningitidis B full length genome sequence and open reading
 XX PT frames are used to detect, treat and prevent Neisserial infections -
 XX PF
 XX PS Claim 7; Appendix A; 692pp; English.
 XX PX
 XX CC The present invention describes the full length genome of
 CC CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC CC sequence were too long to go in a record on its own it was split into 8
 CC CC sequences which overlap each other at the beginning and end of each
 CC CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC CC Neisseria proteins given in AAF21589 to AAF21593, and AAF21593 to
 CC CC AAF21606 represent PCR primers which are used in the exemplification of
 CC CC the present invention. The NMB genome and fragments from it have
 CC CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC CC proteins can be used in compositions for treating or preventing infection
 CC CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
 CC CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC CC bacteria. Computers, computer memory, computer storage medium or computer
 CC CC databases can be used in a search to identify open reading frames (ORFs)
 CC CC or coding sequences within the NMB genome. The DNA sequences provide
 CC CC further opportunities to find antigenic or immunogenic proteins which are
 CC CC more effective in vaccines than the outer membrane proteins currently
 CC CC used.
 XX SQ Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 349980
 Score: 3927.00 Matches: 743
 Percent Similarity: 98.81% Conservative: 7
 Best Local Similarity: 97.89% Mismatches: 8
 Query Match: 97.30% Indels: 2
 DB: 21 Gaps: 0
 US-09-936-377-2 (1-758) x AAF21609 (1-349980)
 QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
 DB 85503 ATGGCAAAACACTCACTCAAAACCAATGTTTATCAATCTTTTATCAACACACCCCTC 85444
 QY 21 LeuAlaGlnAlaHisGlnThrGluGlnSerValGlyLeuGlnThrValThrValValGly 40
 DB 85443 CTCGCCACGCGATGAATGAGCAATCGTGATTTGAAACGCGTCAGCGTCGCGC 85384
 QY 41 LysSerArgProAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 DB 85383 AAAAGCGTCGCGCGCGCAGTCGCGGCTGTTGCACACTTCGACGCGCTTCGACAAAATC 85324
 QY 61 IleSerGlyAspThrLeuArgGlnIleValAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 DB 85323 ATCTCCGGCGATACCTTGGCCAAAGCGTCAACTTGGGCGAGCGCTTTAGACGCGCTA 85264
 QY 81 ProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 DB 85263 CCGGGCATCCAGCTTCGCAATACGCGCGCGCGGTCTGCTCCCGTCATTCGCGGTCAA 85204
 QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 DB 85203 ACAGGCGAGCGGATTAAGTGTGAACCATCAAGCGGAAACAGGCGATATGCGCGATT 85144
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 DB 85143 TCSCCGCATACGCCATTATGTAGATACCGCTTGTGCAACAGGTGCAATCTCGCGC 85084
 QY 141 GlyProValThrLeuLeuThrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 DB 85083 GGGCGGTTACGCTCTTGTACAGCTCGGCAATGTGGCGGGCTGCTCGATGTTCCGAT 85024
 QY 161 GlyIleProGlnLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 DB 85023 GGCATAATCCCGAAAAATGCTGAAACGCGGTATCGGCGCAACTCGGATTCGCTTTG 84964
 QY 181 SerSerGlyAsnLeuGlyLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 DB 84963 AGCAGCGGCAATCTGAAAAACTCAGTCCGCGGCGATCAATATCGGTTTGGGCAAAAC 84904
 QY 201 PheValLeuHisThrGluGlyLeuThrArgLysSerGlyAspThrAlaValProArgTyr 220
 DB 84903 TTTGTATTGACACAGGAAGGCTGTACCGCAATCGGGGATTCGCGGTACCGGCTTAC 84844
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 DB 84843 CGCAATCTGAAACGCTTCCCGACAG - CCACGCGGATTCGCAACGCGCAGCATCGGCT 84785
 QY 241 ValLeuGlyTrpArgLysArgPheThrArgArgThr - TyrSerAspArgArgAspGlnTyr 260
 DB 84784 GTCTTGGTTGGCGAAAAAGTTTATTCGCGGTAGCGGTACAGCGACCGCTCGGCAATA 84725
 QY 260 rGlyLeuProAlaHisSerHisGlnThrAspAspCysHisAlaAspIleIleThrGlnly 280
 DB 84724 TGTCTGCTGCTGCCACAGCCAGCATCATGATGTCACGCGCATCATCTATCGCAAAA 84665
 QY 280 sSerLeuIleAsnLysArgTyrIleGlnLeuThrProHisLeuLeuThrGluAspVa 300
 DB 84664 GAGCTTGATTAACAAACGCTATTATACAGCTTTTATCGGCACCTGTTGACCGAAGAGACAT 84605
 QY 300 LaspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
 DB 84604 CGATTACGACATCCCGGCTTGAGCTGCGGCTTCCACGACGACATATATGACACCGACA 84545
 QY 320 sAlaHisAsnGlyLysProThrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG 340
 DB 84544 CACCCACAGCGCGAGACCGGTGATAGACCTTGGCAACAAACGCTACGAACTCGGTGCCA 84485
 QY 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
 DB 84484 ATGAAGAACACCGTTCCCGGTTTGAAGCCCTGCGCGTACCTGAAACCGCAAGACATPA 84425
 QY 360 rHisHisAspGlyLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 DB 84424 CCGCCACGAGAAAAAGCAGCGGATCGATCGCAAAACTTTTTTAAACAAACCAAA 84365
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerThrGlyValG 400
 DB 84364 CGCCCGCATCGATGTCGCGCACCAACCCATAGTCTGTAAGAGCGGCGCTGGGGGTGCA 84305
 QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
 DB 84304 ATATTTACAAACAAATCCAGTGCTTTATCTGCCATATCCGAGGCGGTAAACACCGAT 84245

Db 61 CTGCCCCAAGCGCATGAACACTGAGCAATCGGTGGATTTGGAAACGGTCAGCGCTCGTCGCG 120
 Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 121 AAAAGCCGTCCGCGGCCACAGTCGCGGCTGTGTGCACATTCAGCCCGCTCCGACAAATC 180
 Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 181 ATCTCCGGCGATACCTTGCGCCAAAAGCCGTCACACTTGGGCGACGCTTGAAGCGCGTA 240
 Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 241 CCGGCGATCCAGCTTCGCAATACGCGCGCGCGCTGTGCTCCGCTCATTCGCGGTCAA 300
 Qy 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 301 ACAGCGAGCGGATTAAGTGTGAACCATCACGCGGAAACAGCGGATATGCGCGATTTT 360
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 361 TCGCCCGATCAGCCCATATGTGTAGATACCGCTTGTGCGAACAGGTGCAATCTCGGCG 420
 Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValalaasp 160
 Db 421 GGGCCGGTTACGCTCTTGTACAGCTCGGCAATGTGGCGGGCTGGTGGATGTTGCCGAT 480
 Qy 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 481 GCGAAATCCCGCAAAATGCCTGAAACGCGGTATCGGGGAACTCGGATTCGTTTG 540
 Qy 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 541 AGCAGCGGCAATCTGGAATAATCTACGCTCCGCGCGCATCAATATCGGTTGGGCAAAAC 600
 Qy 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 601 TTGTGATTGCACACGGAAGGCTGTACCGCAATCGGGGATTAACGCGTACCGGTTAC 660
 Qy 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 661 CGCAATCTGAACGGCTGCCCGACAG-CCACGCCGATTCGCAACCGGCGAGCATCGGGCT 719
 Qy 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
 Db 720 GTCCTGGGTGGCAAAAGGTTTATCGCGGTAGCGTACAGCGACCGTCGCGACCAATA 779
 Qy 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
 Db 780 TGGTCTGCTGCCACAGCACAGATAGATGATTGCCACGCCGACATCATCTGCGCAAA 839
 Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysAspVa 300
 Db 840 GAGCTTGATTAAACAACGCTATTTACGCTTTATCCGACCTGTTGACCGAAGAAGACAT 899
 Qy 300 LasPyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
 Db 900 CGATTAACGACATCCGGGCTTGAGCTGGGCTTCCACGACGATATGACACGCGACA 959
 Qy 320 sAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG 340
 Db 960 CACCCACAGCGCGACAGCGTGGATAGACCTGGCGCAACAAACGCTACGAACTCCGTGCCGA 1019
 Qy 340 uTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
 Db 1020 ATGGAAGCAACCGTTCCCGGTTTTGAAGCCCTGGCGGTACACCTGAAACCGCAACGACTA 1079
 Qy 360 rHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 Db 1080 CCGCCACAGCGAAAAGCGGCGATGCGATCGTGAATACTTTTAAACAACCAACGCAAAA 1139
 Qy 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValG 400
 Db 1140 CGCCCGCATCGAGTTGGCGCCCAACCCCATAGGTGCTCTGAAAGGCGAGCTGGGGGTGCA 1199

Qy 400 mTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
 Db 1200 ATATTTTCAACAACAAATTCAGTGCCTTTATCTGCCATATCCGAAGCGGTAAACAACCGAT 1259
 Qy 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 Db 1260 GCTGCTTGACAAACAAAGTCAACATTCAGCTTTTTCGGTGTAGAACAAGCAACTGGGA 1319
 Qy 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 Db 1320 CAACCTTCAGCTTGAAAGAGGCGTACGCGTGAAAAACAAAAAGGCTCCATTCAGTACGA 1379
 Qy 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyVala 480
 Db 1380 CAAGCATTTGATTCGGGAAAACTACTACACACCCCTCGCCGACCTCGGCGCGCA 1439
 Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHis 500
 Db 1440 CCGCCAAACCCCGCTCATTCGCACCTTTTCGGGCAACTGGTATTTTCAGCCCAACACAA 1499
 Qy 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla 520
 Db 1500 ACTCAGCTGACCGCTCCCATCAGAAACGCGCTCGCTCAACGCAAGAGCTGTACGCACA 1559
 Qy 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 Db 1560 CGGCAAAACAGCTCGCCACCAACACCTTTGAAAGTCGGCAACAAACACCTCAACAAGAGCG 1619
 Qy 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
 Db 1620 TTCCAAACAATATCGAACTCGGCTCGGCTAGAAAGCGACCGCTGCGCAATCAATCTGCGC 1679
 Qy 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG 580
 Db 1680 ACTTACCGCAACCGCTTCGGTACTACTTTTACGCCCAACCTTAAACGACGCGACGCGG 1739
 Qy 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 Db 1740 CCCCATAATCCATCGAAGACGACGAGCAATGAAGCTCGTGCCTACCAACCAATCCGCGC 1799
 Qy 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG 620
 Db 1800 CGACTTCTACGCGCGGAAGCGAAATCTACTTCAACCGACACCGCTACCGCATCGG 1859
 Qy 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG 640
 Db 1860 COTTTCCGGCGCACTATGTACGAGCGCGTCTGAAAAACCTGCTTCCCTACCGCGGAGAGA 1919
 Qy 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 Db 1920 AGATSCCTACGCAACCGCTCTTTCATCGCAGGACGACCAAAATGCCCCCGTGTTC 1979
 Qy 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
 Db 1980 GGTCTCGCGCTCGGCTTCCACCTGAAAGCGCTCGCTGACCGACCGCTATCGATGCAATTT 2039
 Qy 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG 700
 Db 2040 GGACTACTACCGGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGCACCGCCCGG 2099
 Qy 700 yHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAs 720
 Db 2100 ACACCATATGCTCAACCTCGGCGCAACACTACCGCGCAATACCGCTATGCGGAGTGGNA 2159
 Qy 720 nTrrTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 2160 TTGTACGTCAAGCGCAACCTGCTCAACAAATCCGTTTACGCCCAACGACGAGCTTCT 2219
 Qy 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
 Db 2220 CTCGTATACGCCGCAATATGGGCGCGAGCTTTTACCGCGCGCGTGAACGTGAAGTTT 2274

RESULT 7	DB:	21	Gaps:	0
AAA81490/c	US-09-936-377-2 (1-758) x AAA81490 (1-1437668)			
ID AAA81490 standard; DNA; 1437668 BP.				
AC AAA81490;				
XX 04-DEC-2000 (first entry)				
XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.				
DE				
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;				
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;				
KW Meningococcus B; MenB; ds.				
XX Neisseria meningitidis.				
OS				
XX WO200022430-A2.				
FN				
XX 20-APR-2000.				
PD				
XX 08-OCT-1999; 99WO-US23573.				
PF				
XX 09-OCT-1999; 98US-0103794.				
PR				
XX 30-APR-1999; 99US-0132068.				
XX (CHIR) CHIRON CORP.				
PA				
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;				
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V;				
PI Rappuoli R, Pizza M;				
XX WPI; 2000-318079/27.				
DR				
XX Isolated nucleotide sequences of Neisseria meningitidis which can be				
PT used in the diagnosis and treatment of N. meningitidis infection and				
PT other Neisserial infections, for example, N.gonorrhoea -				
XX Claim 7; Page 866-1272; 1760pp; English.				
PS				
XX The present invention describes methods of obtaining immunogenic				
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414				
CC represent specifically claimed Neisseria meningitidis genomic DNA				
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent				
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to				
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the				
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to				
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF				
CC sequences, which are all used in the exemplification of the present				
CC invention. The nucleic acid sequences, protein sequences, and antibodies				
CC against them, can be used in the manufacture of a composition. The				
CC composition can be used as a medicament (or in the manufacture of a				
CC medicament) for treating, preventing or diagnosing infection due to				
CC Neisserial bacteria. For example, some of the identified proteins could				
CC be components of vaccines against Meningococcus B; against all serotypes;				
CC and/or against all pathogenic Neisseriae. Identification of sequences				
CC from the bacterium will also facilitate production of biological probes,				
CC particularly organism-specific probes. Attempts to make efficacious				
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.				
CC Multivalent vaccines have also been tried but none have successfully				
CC overcome antigenic variability. The provision of further, complete				
CC sequences may provide an opportunity to identify secreted or surface				
CC exposed proteins that may be presumed targets for the immune system and				
CC which are not antigenically variable or at least more conserved than				
CC other more variable regions.				
XX				
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;				
Alignment Scores:				
Pred. No.:	0	Length:	1437668	
Score:	3922.00	Matches:	742	
Percent Similarity:	98.68%	Conservative:	7	
Best Local Similarity:	97.76%	Mismatches:	9	
Query Match:	97.18%	Indels:	2	

```

Db 984483 ATGGAGCAACCGTTCCCGGTTTGAAGCCCTGGCGGTACACCTGAACCGCAACGACTA 984424
Qy 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
Db 984423 CGCCACGACGAAAGAGGAGCGATGCGTGTGAAACCTTTTAAACACCAACGCAAAA 984364
Qy 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValG 400
Db 984363 CGCCCGCATCGAGTTGCGGCACCAACCCCATAGTGTCTGTGAAGAGCAGCTGGGGCGTGCA 984304
Qy 400 nTyrrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
Db 984303 ATATTTACACAAATAATCCAGTCTTATCTGCCATATCCGAAGCGGTAAACACCGAT 984244
Qy 420 tLeuLeuAspAsnLysValGlnHisTyrrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
Db 984243 GCTGCTTGACAAACAAAGTGCACATTTACAGCTTTTTCGGTGTAGACACGCAAACTGGGA 984184
Qy 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrrAs 460
Db 984183 CAACCTTACACCTTTGAAGAGCGGTACGCGTGGAAAAACAAGAGCTTCATTTCAGTACGA 984124
Qy 460 pLysAlaLeuIleAspArgGluAsnTyrrTyrrLysGlnProLeuProAspLeuGlyAlaHi 480
Db 984123 CAAAGCATTTGATTGATCGGGAAACCTACTACACACACCCCTCCCGCACTCGGCGCGCA 984064
Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrrPheThrProGlnHisLys 500
Db 984063 CCGCCAAACCGCCCGCTCATTCGCACCTTCGGCAACTGTTATTTACGCCCAACACACAA 984004
Qy 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrrAlaHi 520
Db 984003 ACTCAGCTCAGCGCTCCCTCCATCAGGAACGCTCGCGTCAACGCAAGAGCTGTACGACA 983944
Qy 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
Db 983943 CGGCAACACGTCGCCACCAACCTTTGAAGTGGCAACAAACACTCAACAAAGAGCGG 983884
Qy 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrrGluGlyAspArgTrpGlnTyrrAsnLeuAl 560
Db 983883 TTCCACAAATATCGAATCGCGCTGGGCTACGAAGCGCGCTGCGCAATACATCTGCGC 983824
Qy 560 aLeuTyrrArgAsnArgPheGlyAsnTyrrIleTyrrAlaGlnThrLeuAsnAspGlyArg 580
Db 983823 ACTCTACCGAACCGCTTCGCTAACTACATTTACGCCCAACCTTAAACGACGACGCGG 983764
Qy 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrrAsnGlnSerGlyAl 600
Db 983763 CCCCAATCCATCGAAGACGACAGCGAAATGAAGCTCGTGGCTACACCAATCCGCGCGC 983704
Qy 600 aAspPheTyrrGlyAlaGluGlyGluIleTyrrPheLysProThrProArgTyrrArgIleG 620
Db 983703 CGACTTCTACGCGCGGAAGCGGAAATCTACTTCAACACCGACACCCGCTACCGCATCGG 983644
Qy 620 yValSerGlyAspTyrrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG 640
Db 983643 CGTTTCCGCGACATATGTACAGAGCGCTGTGAAGAAACCTGCCCTCCCTACCCGCGAGAG 983584
Qy 640 uAspAlaTyrrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
Db 983583 AGATGCTTACGCGCAACCGTCTCTTTCATCGCACAGAGACCAAAATGCCCGCGGTGTTCC 983524
Qy 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
Db 983523 GCGTCGCGCGCTTCGCTTCCACTTGAAAGCCCTCGCTGACCGACCGTATCGATCGCAATTT 983464
Qy 680 uAspTyrrTyrrArgValPheAlaGlnAsnLysLeuAlaArgTyrrGluThrArgThrProG 700
Db 983463 GGACTACTACCGCGTGTTCGCCCAACCAAACTCGCCCGCTACGAAACGCGCACGCCCGG 983404
Qy 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrrArgAsnThrArgTyrrGlyLutrpAs 720

```

```

Db 983403 ACACCATATGCTCAACCTCGCGCAAACTACCGCCCAATACCGCTATGCGGAGTGGAA 983344
Qy 720 nTrpTyrrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrrAlaHisSerSerPheLe 740
Db 983343 TTGTTAGTCTCAAAAGCCGACAACTGTCTCAACCAATCCGTTTACGCCACACGAGCTTTCT 983284
Qy 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 983283 CTCTGATACCCCGCAATGCGCCGCGACCTTACCGCGCGGTGAACGTGAAGTTT 983229

RESULT 8
AAZ53328
ID AAZ53328 standard; DNA; 2277 BP.
XX
AC AAZ53328;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 149 partial DNA sequence SEQ ID NO:605.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 03-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAV74566.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 424; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 2277 BP; 598 A; 721 C; 547 G; 411 T; 0 other;

```

Alignment Scores:

Pred. No.:	0	Length:	2277
Score:	3920.00	Matches:	743
Percent Similarity:	98.55%	Conservative:	5
Best Local Similarity:	97.88%	Mismatches:	10
Query Match:	97.13%	Indels:	2
DB:	21	Gaps:	0
US-09-936-377-2 (1-758) x AAZ53328 (1-2277)			
Qy	1	MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu	20
Db	1	ATGGCACAACTACACTCAAAACCCATTGTTTATCAATTCCTTTAATCAACACACCCCTC	60
Qy	21	LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValValGly	40
Db	61	CTCTCCCAAGCGCATGGAACCTGAGCAATCAGTGGGCTTGGAACCGTACGCGTCGTCGC	120
Qy	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
Db	121	AAAGCCGTCGCGCGCCACATTCGGGGCTGTGCACTCTACCGCCCTCCGACAAATC	180
Qy	61	IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal	80
Db	181	ATCAGCGCGACACCTTGCACAAAAGCCGTCACTTGGTGATGCTTTAGACGCGTA	240
Qy	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln	100
Db	241	CCGGCATTTCAATGCTCGCAATACGCGCGCGGCATCCGCTCCCGTTATTCGCGTCAA	300
Qy	101	ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe	120
Db	301	ACAGCAGACGGATTAAAGTGTGAACCATCACGCGCAACCGGGCGACATGCGGACCTC	360
Qy	121	SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg	140
Db	361	TCTCCAGACCATGCAATCATGTGTGACAGCGCTTGTGCAACAGGTCGAAATCCTCGCG	420
Qy	141	GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp	160
Db	421	GGTCGGGTACGCTCTTGTACAGCTCGGCAATGTGGGGGCTGGTGCATGTTGCCGAT	480
Qy	161	GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu	180
Db	481	GGCAAAATCCCGAAAAATGCTGAAACGCGGTATCGGGGCACTCGATGTCGTTTG	540
Qy	181	SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn	200
Db	541	AGCAGCGCAATCTGGAAAAATCACTCGTCGGCGCATCAATATCGTTTGGGCAAAAC	600
Qy	201	PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr	220
Db	601	TTTGTATTGCACCGAAGGGCTGTACCGCAATCGGGGATACGCGCATCGCGGTAC	660
Qy	221	ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla	240
Db	661	CGCAATCTGAAACGCTGCCCGACAG-CCACGCCGATTCGCAACGGGACGATCGGCT	719
Qy	241	ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr	260
Db	720	GTCTTGGGTTCGGCAAAAGGCTTTATCGCGCAGCATACACGACCGTCGCGACCAATA	779
Qy	260	rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnIly	280
Db	780	TGGTCTGCTGCCACACGCCAGCAATACATGATTGCCACGCCGACATCATCTGGCAAAA	839
Qy	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa	300
Db	840	GAGTTTGTATAACAAACGCTATTTCAGCTTTATCCGACCTTGTGACCGAAGACAT	899
Qy	300	1AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
Db	900	CGATTACACAATCCGGCTTGACTGGGGCTTTACGACGACGATGATGCACACGCCCA	959

Db 2040 GGAAGTACTACCGCGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGCACGCCCG 2099
 QY 700 YHSHISMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrpAs 720
 Db 2100 ACACCATATGCTCAACTCCGCCAAACTACCGCCGCATACCGCTATGGCGAGTGGAA 2159
 QY 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 2160 TTGGTACGTCAAAAGCGGACAACTGCTCAACCAATCCGTTTACGCCCCACAGAGCTTCT 2219
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2220 CTCTGATACGCCGCAATGGCCGCGCTTTACCGCGCGGTGAACGTGAAGTTT 2274

RESULT 9
 AAZ53326
 ID AAZ53326 standard; DNA; 2277 BP.
 AC AAZ53326;
 DT 21-MAR-2000 (first entry)
 XX Neisseria gonorrhoeae ORF 149 partial DNA sequence SEQ ID NO:601.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 OS Neisseria gonorrhoeae.
 PN WO9957280-A2.
 XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR P-P8DB; AAY74564.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics

PS Claim 7; Page 421; 1453pp; English.
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.
 XX Sequence 2277 BP; 599 A; 726 C; 544 G; 408 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 2277
 Score: 3638.00 Matches: 726
 Percent Similarity: 96.84% Conservative: 9
 Best Local Similarity: 95.65% Mismatches: 23
 Query Match: 95.09% Indels: 2
 DB: 21 Gaps: 0
 US-09-936-377-2 (1-758) x AAZ53326 (1-2277)
 QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleThrProLeu 20
 Db 1 ATGGCCAAATCACACTCAACCCATGTTTATCAATTCCTTTAATCAACACACCCCTC 60
 QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 Db 61 CTCGCCAAAGCGCATGAAACTGAGCAATCGTGGGCTTGGAAACGCTCAGCGTCGTCGC 120
 QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 121 AAAAGCGTCCGCGCGGACCTTCGGGGCTGCTGCACACTTCGACCGCTCCGCAAAATC 180
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 181 ATCTCGGCGATACCTTGGCCCAAAAGCGCTCACTTGGCGACCTTTGGACGGCGTA 240
 QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 241 CCGGGCATCCACGCTTCGCAATACGGCGCGGCGCATCCGCTCCCTTATTTCGCGTCAA 300
 QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 301 ACGGCGAGACGGATTAAAGTATTGAACCATCACGGGAAACGGCGCATATGGCGGACTTT 360
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 361 TCTCCGATCAGCCCATATATGTTAGTATACCGCTTGTCCAAACAGGTTTGAATCTCGCGC 420
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 421 GGGCGGTTTACGCTCTTGTACAGCTCGGCAATGTGGCGGGCTGCTCGATGTTGCGCAT 480
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 481 GGAAAAATCCCCGAAAAAATGCTGAAACGGCGTATCGGGCGGAAACCGGATTCGCTTTG 540
 QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 541 AGCAGGGCAATTTAGAAAAACTGACATCCGAGGATCAATATCGGACTGGGCAAAAAC 600
 QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 601 TTCGTGCTCATACCGAAGCGCTTGTACGCAATCGGGCGATTACGCGCTACCGCGTTAC 660
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 661 CCAATCTGAAACGCTGCCCGACAG-CCATCCCGATTTCGCAACCGGCGAGCATCGGGCT 719
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
 Db 720 GTCTTGGTGGGCGAAAGGCTTTATGGCGCAGCATACAGCGACCGTCCGCGACCGCTA 779
 QY 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnly 280
 Db 780 CGGCGCTGCTGCCACAGCCACGAATACGATTGCGCAGCGCGCATCATCTGGCAAAA 839
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysPhe 300
 Db 840 GAGTTTGATCAACAAACCGCTATTTCGAGCTTTATCCGCACTTGTGTGACCGAAGACAT 899

```

QY 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
Db :|||||
QY 900 CGATTACGACAAATCCGGGCTTGAGCTGGCGCTTCACGACGGCGACGGTGCAACGCACA 959
Db :|||||
QY 320 sAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAlaG1 340
Db :|||||
QY 960 CACCCACAAACGCAACCGTGATAGACTCGGCAACAAACGCTACGAACCTCGGCGCGCA 1019
Db :|||||
QY 340 uTriPysGlnPropheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
Db :|||||
QY 1020 ATGGAAGCAGCCATTCGCCGGTTTTGAAGCCCTGGCGGTACATCTGAACCGCAATGACTA 1079
Db :|||||
QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
Db :|||||
QY 1080 CCACCACAGCAAGAAAGCAGCGGATGAGTGAAGAACTTCCTTCAACAAACAAACACACAA 1139
Db :|||||
QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrPglValG1 400
Db :|||||
QY 1140 CGCCCGTATCGAGTTGGCCACCAACCCCATAGCCGCTCTGAAGGCGAGCTGGGGGTGCA 1199
Db :|||||
QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
Db :|||||
QY 1200 ATATTGGGACAAAATCCAGCGCGCTTTCCGCCATTTCCCGAAACCGTCCAAACAAACCGAT 1259
Db :|||||
QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
Db :|||||
QY 1260 GTTGATTGACAAACATGTCCGCCATTACAGCTTTTCGGTGAGAACAGGCAAAATGGGA 1319
Db :|||||
QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db :|||||
QY 1320 CAACITTCAGCTTGAAAGCGCGGTACGGGTGAAACAAACAAAGCCTCCATCCGGTACGA 1379
Db :|||||
QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
Db :|||||
QY 1380 CAAGCATTTGATGATCGAGAAATACTACAAACAGCCCGCTCGCGCGCGCA 1439
Db :|||||
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLys 500
Db :|||||
QY 1440 CGCCCAACCGCCGCTGTTGCGACTTTCGGGCAACTGGTATTTCACGCCACACACAA 1499
Db :|||||
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db :|||||
QY 1500 ACTCAGCTGACCGGCTCCCATCAGAACCGCTCGGTCAACGCAAGAACTGTACGCACA 1559
Db :|||||
QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
Db :|||||
QY 1560 CGGCAAGCAGCGTCGCCAACACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG 1619
Db :|||||
QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
Db :|||||
QY 1620 TTCCAAACATATCGAACTCGCGCTGGGTACGAAGGCGACCGCTGGCAATACATCTGCG 1679
Db :|||||
QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG1 580
Db :|||||
QY 1680 AGCCTACGCGCAACCGATTTCGGCACTACATTTACGCCCAACCTTAAACGACGCGACGGG 1739
Db :|||||
QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db :|||||
QY 1740 CCCCAATTCATCGAAGACGACGAGCAATGAAGCTCGTGGCTACAAACCAATCCGGTGC 1799
Db :|||||
QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG1 620
Db :|||||
QY 1800 CGACTTCTACGGCGCGGAAGCGGAATCTACTTCAACCGACACCGCTACCGCATCGG 1859
Db :|||||
QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG1 640
Db :|||||
QY 1860 TGTTTTCCGGCGCACTATGTACGAGGCGGTCTGAAAAACCTCGCGTCCCTACCGCGCAGGA 1919
Db :|||||
QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
Db :|||||
QY 1920 AGATCCCTACGGCAAAACGTCCCTTCATCGCAAGCCGACCAAAACGCCGCCGATTC 1979

```

```

QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaHisLe 680
Db :|||||
QY 1980 GGCCTGGCGCTCGGCTTCCACCTGAAAACCTCGTAAACCGGCTATCGATGCCAATTT 2039
Db :|||||
QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG1 700
Db :|||||
QY 2040 GGACTACTACCGGTGTTGCGCCAAACAACTCGCCGCTACGAACCGGTACGCCCGG 2099
Db :|||||
QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAs 720
Db :|||||
QY 2100 ACACCATATGCTCAACCTCGGTGCAAACTACCGCGCAATACGCGTATGGCGAGTGAA 2159
Db :|||||
QY 720 nTriPyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
Db :|||||
QY 2160 TTGGTACGTCAAGCGGCAACCTGCTCAACCAATCCGTTTACGCCCCACAGCAGCTTCCT 2219
Db :|||||
QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db :|||||
QY 2220 CTCTGATACCGCGCAATGGGCGGCAAGCTTTACCGCGCGGTAAACGTGAAGTTT 2274
Db :|||||
RESULT 10
AAZ53318
ID AAZ53318 standard; DNA; 2208 BP.
XX
AC AAZ53318;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 147 partial DNA sequence SEQ ID NO:585.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY74556.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 412-413; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the

```

CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 2208 BP; 579 A; 689 C; 540 G; 400 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 2208
Score: 382.00 Matches: 718
Percent Similarity: 98.78% Conservative: 8
Best Local Similarity: 97.69% Mismatches: 8
Query Match: 94.45% Indels: 2
DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53318 (1-2208)

QY 25 HisGluThrGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
DB 4 CATAAACTAGCAATCGGTGGATTTGGAAACGGTCAGCGTCGTGGCAAAAGCGGTCCG 63
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
DB 64 CGGCCACGTCGGGGCTGTTGGACACTTCGACCGCTCCGACAAATCATCTCCGGCGAT 123
QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
DB 124 ACCTTGGCCAAAGACCGTCACTTGGCGACGCTTACAGCGGCTACCGGCATCCAC 183
QY 85 AlaSerGlnTyrglyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
DB 184 GCTTCGCAATACCGCGCGCGCGCTGTGCTCCGCTCATTCGCGTCAAAACAGGCGCGG 243
QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
DB 244 ATTAAAGTGTGAACCATACCGCGGAACAGGGCGATGCGGATTTTCGCCCGGATCAC 303
QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValIleLeuArgGlyProValThr 144
DB 304 GCCATTATGTAGATACCGCTTGTGCAACAGGTGCAATCTCGCGCGCGGTACG 363
QY 145 LeuLeuTyrsSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
DB 364 CTCTTGACAGCTCGGGCAATGTGGCGGGCTGTGATGTTGCCGATGGCAAAATCCCC 423
QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
DB 424 GAAAAATGCCTGAAACCGCGTATCGGGCAACTCGGATGCGTTTGACGCGCGCAAT 483
QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
DB 484 CTGGAAAAACTCACGTCGCGCGCATCAATATCGGTTTGGCAAAATTTGTATTGCAC 543
QY 205 ThrGluGlyLeuTyrglySerGlyAspTyrglyAlaValProArgTyrglyArgAsnLeuLys 224
DB 544 ACGGAAGGGCTGTACCGCAAAATCGGGGATTTACGCGGTATCCGCGTATCTGAAA 603
QY 225 ArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyr 244
DB 604 CGCCTGCCGACAG-CCACGCGGATTCGCAACAGCGGCGAGCATCGGCTGTCTTGGGTGG 662
QY 245 ArgLysArgPheTyrglyArgThr-TyrSerAspArgAspGlnTyrglyLeuProAl 264
DB 663 CGAAAAAGGTTTTATCGGTAGCTAGCTACAGCGACCGCTCGCGACCAATATGCTGCTGC 722
QY 264 ahisSerHisGluTyrglyAspCysHisAlaAspIleIleTyrGlnLysSerIleLeuAs 284
DB 723 CCACGCCACGATACGATGATTGCCACCGCGACATCATCTGGCAAAAGAGCTTGATTAA 782
QY 284 nLysArgTyrglyLeuGlnLeuTyrglyProHisLeuLeuThrGluGluAspValAspTyrAspAs 304

DB 783 CAAACGCTATTATCAGCTTTATCCGACCTTTGTCACCGAAGACATCGATTACGACAA 842
QY 304 nProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGly 324
DB 843 TCCGGGCTTGAAGTGGCGCTTCCACGACGACGATATATGTCACGACACACACACGCGG 902
QY 324 YLysProTyrIleAspLeuArgAsnLysArgTyrglyGluLeuArgAlaGluTyrglyGlnPr 344
DB 903 CAGACCGTGGATAGACTCGGCAACAAACGCTACGAACTCGTGCCGATATGGAAGCAACC 962
QY 344 oPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrglyHisAspGly 364
DB 963 GTTCCCGGCTTTGAAAGCCCTGGCGGTACACTTGAACCGCAACGACTTACCGCCACGACGA 1022
QY 364 uLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGly 384
DB 1023 AAAAGCAGGCGATGCGAGTCGAAACCTTTTAAACCAACCAACCAACCAACCGCATCGA 1082
QY 384 uLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrglyValGlnTyrglyGlyGly 404
DB 1083 GTTGGCCCAACCAACCATAGTCTGTGAAGGCGAGCTGGGGGTGCAATATTTTACAAACA 1142
QY 404 nLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
DB 1143 AAAATCCAGTCTTTATCTGCCATATCCGAAGCGGTTAAACCAACCGATGCTGCTTGACAA 1202
QY 424 nLysValGlnHisTyrglySerPheGlyValGluGlnAlaAsnTyrglyAspAsnPheThrLe 444
DB 1203 CAAAGTGCACATTTACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGACAACTTCACGCT 1262
QY 444 uGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrglyAspLysAlaLeuIle 464
DB 1263 TGAAGGAGGCGTACGCGTGGAAAAACAAAGAGCGCTCCATTCAGTACGACAAAGCATTGAT 1322
QY 464 eAspArgGluAsnTyrglyLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAl 484
DB 1323 TGATCGGAAAACTACTACAAACCCCTCTGCGGACCTCGGCGCGCACCGCAACCGC 1382
QY 484 aArgSerPheAlaLeuSerGlyAsnTyrglyPheThrProGlnHisLysLeuSerLeuTh 504
DB 1383 CCCTCATTCGCACTTTCGGGCAACTTGTATTTTCAGCCCAACACACAACTTCAGCGCTGAC 1442
QY 504 rAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrglyAlaHisGlyLysHisVa 524
DB 1443 CGCCTCCCATCAGGAACGCTCGCGTCAACGACAGAGCTGTACGACACGCGCAAAACACGT 1502
QY 524 lAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIle 544
DB 1503 CGCCACCAACACCTTTGAAAGTCGGCAACAAACACCTTCAACAAAGAGCGTTCCCAACAATAT 1562
QY 544 eGluLeuAlaLeuGlyTyrglyAspArgTyrGlnTyrglyAsnLeuAlaLeuTyrglyArgAs 564
DB 1563 CGAATCTGGCGTGGGTACGAGGCGACCGCTGGCAATCAATCTGGCACTTCTACCGCAA 1622
QY 564 nArgPheGlyAsnTyrglyTyrglyAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIle 584
DB 1623 CGCTTCGGTAACTACATTTACGCCCAACCTTAAACGACGCGCGCGGCCCAAAATCCAT 1682
QY 584 eGluAspAspSerGluMetLysLeuValArgTyrglyAsnGlnSerGlyAlaAspPheTyrgly 604
DB 1683 CGAAGACGACGAGAAATGAGCTGTGCGCTACAAACCAATCCGCGCGCGACTTCTACCG 1742
QY 604 yAlaGluGlyGluIleTyrglyPheLysProThrProArgTyrglyArgIleGlyValSerGlyAs 624
DB 1743 CGCGAAGCGGAAATCTACTTCAAAACCGACACCGCGCTACCGCATCGGCGTTTCGCGCGA 1802
QY 624 pTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrgly 644
DB 1803 CTATGTAGGAGCGCGTCTGAAAAACCTGCTTCCCTCCCGCGACAGAGAGATGCTTACCG 1862
QY 644 yAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLys 664
DB 1863 CAACCGTCTTTTCATCGCACAGGACGACAAATGCCCGCGCTGTTCGCGCTGCGCGCT 1922

QY 664 uclypheHisLeuLysLeuSerLeuThrAspArgLeuAspAlaLeuAspTyrTyrAr 684
 Db 1923 CGGCTTCACCTGGAAGCCTCGCTGACGACCGATCGATGCAATTTGGACTACTACCG 1982
 QY 684 gValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLe 704
 Db 1983 CGTGTTCGCCCAAACTCGCCGCTTACGAAACGGCGACGCCCGGACCATATGCT 2042
 QY 704 uLenLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluThrAsnTyrTyrVally 724
 Db 2043 CAACCTCGCGCGCAAACTACCGCGCAATACCGGCTATGCGGAGTGGAAATTGGTACGTCAA 2102
 QY 724 salAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPr 744
 Db 2103 AGCGGACACCTGCTCAACCAATCCGTTACGCCACACAGCAGCTTCTCTGTATACGCC 2162
 QY 744 oGlnMetGlyArgSerPheThrGlyGlyValLeuValLysPhe 758
 Db 2163 GCAATGGGCGCAGCTTTACCGGCGCGTGAACGTGAAGTTT 2205

RESULT 11

AAA81500/c
 ID AAA81500 standard; DNA; 22863 BP.

XX
 AC AAA81500;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_47 SEQ ID NO:47.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI; 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7; Page 1310-1317; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisariae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX
 SQ Sequence 22863 BP; 5657 A; 5469 C; 5921 G; 5816 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 22863
 Score: 3812.00 Matches: 718
 Percent Similarity: 98.78% Conservatives: 8
 Best Local Similarity: 97.69% Mismatches: 8
 Query Match: 94.45% Indels: 2
 DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAA81500 (1-22863)

QY 25 HisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
 Db 22860 CATAAACTGAGCAATCGTGGATTTCGAAACCGTCAGCGTCGCGCAAAAGCGCTCG 22801
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
 Db 22800 CGCGCCAGCTCGGGCTGTTCACACTTCGACCCCTCCGACAAATCACTCCGGCGAT 22741
 QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
 Db 22740 ACCTTGGCCAAAAGCGCTCAACTTGGCGACGCTTTAGACGCGTACCGGCATCCAC 22681
 QY 85 AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
 Db 22680 GCTTCGCAATACGGCGCGCGCTCTGCTCCCGTCATTCGCGGTCAACACAGCGCGGG 22621
 QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 Db 22620 ATTAAGTGTGTAACCATCACGCGCAACAGCGCATATGGCGGATTTTCGCCCGATCAC 22561
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGlyProValThr 144
 Db 22560 GCCATTATGTAGATACCGCTTGTGCAACAGGTTCGAAATCTCTGCGCGCGCGTTACG 22501
 QY 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 Db 22500 CTCITGTACGCTCGGCAATGTGCGGGGCTGTGTGATGTTGCGATGGCAAAATCCCC 22441
 QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
 Db 22440 GAAAAAATGCTGAAAAACGGCGTATCGGCGCAATCGGATTCGCTTTGAGCAGCGGCAAT 22381
 QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
 Db 22380 CTGAAAAAATCTCACGTCGCGCGGATCAATATCGGCTTTGGGCAAAATCTTTGATGTGAC 22321
 QY 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys 224
 Db 22320 ACGGAAGGGTGTACCGCAATCGGGGATTTACCGGTACCGGTACCGCAATCTGAA 22261
 QY 225 ArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyr 244
 Db 22260 CGCCTCGCCGACAG-CCACGCGGATTCGCAACAGCGGACGATCGGCTGTCTTGGGTGG 22202
 QY 245 ArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyrGlyLeuProAl 264
 Db 22201 CGAAAAAGGTTTATTCGGCGTAGCGTACGACCGCTCGGACCAATATGTCGCTGCG 22142

Alignment Scores:

```

Pred. No.: 0 Length: 2184
Score: 3706.00 Matches: 698
Percent Similarity: 96.98% Conservative: 9
Best Local Similarity: 95.75% Mismatches: 21
Query Match: 91.82% Indels: 2
DB: 25 Gaps: 0

US-09-936-377-2 (1-758) x AB241893 (1-2184)

QY 31 ValGlyLeuGluThrValThrValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db 1 GTGGGCTTGAAACGGTCAGCGTCGTCGCGCAAAAGCGGTCGCGCGCGACTTCGGGGCTG 60
QY 51 LeuHisThrSerThrAlaSerAspLysIlelleSerGlyAspThrLeuArgGlnLysAla 70
Db 61 CTGGCACACTTCACCGCGCTCCGACAAATCATCTCCGGCGATACTTTGGCGCCAAAAGCC 120
QY 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGly 90
Db 121 GTCACTTGGCGAGCGCTTTGGACGGCGTACCGGGCATCCAGCTTCGCAATACGGCGCG 180
QY 91 GlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHis 110
Db 181 GCGCGATCCGCTCCCGTTATTCGGGTCAACCGGCGAGCGGATTAAAGTATTGAACCAT 240
QY 111 HisGlyGluThrClyAspMetAlaaspPheSerProAspHisAlaIleMetValAspThr 130
Db 241 CACGGCGAAACGGCGGATATGGCGACTTTTCTCCGATCACGCCATTATGTAGTATACC 300
QY 131 AlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGly 150
Db 301 GCCTTGTCCACACAGTTGAATCTTCGCGGGCGGTTACGCTCTGTACAGCTCGGGC 360
QY 151 AsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
Db 361 AATGTGGCGGGCTGTGTCGATGTGCGGATGAAATAATCCCGGAAAAATGCTGAAAC 420
QY 171 GlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer 190
Db 421 GCGGTATCGGGCGAGCGCGATGCGTTTGAGCGCGGCAATTTAGAAAAATGACATCC 480
QY 191 GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg 210
Db 481 GCAGGCATCAATATCGACTGGGCAAAACTTCGTGCTGCATACCGAAGCTTTGACCGC 540
QY 211 LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro 230
Db 541 AAATCGGGCGATTAGCGCGTACCGGTTTACCGCAATCTGAAACGCGCTGCCCGACAG-CCA 599
QY 231 ArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArg 250
Db 600 TGCCGATTTCGCAACCGGCGAGCATCGGCTGTCTTGGGTGGCGGAAAAAGGCTTTATCGG 659
QY 251 ArgThr-TyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAs 270
Db 660 CGCAGCATACACGACCGTTCGCGACCGTACCGCTGCTGCTGCCACAGCCACGATACGA 719
QY 270 pAspCyHisAlaAspIlelleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLe 290
Db 720 TGAATTGCCACGCGCATCATCTGGCAAAAGAGTTTGATCAACAAACGCTATTTCAGCT 779
QY 290 uTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGl 310
Db 780 TTATCCGACCTTTGTACCGGAAGAGACATCGATTACGACAAATCCGGGCTTGAGCTGGG 839
QY 310 yPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAspLe 330
Db 840 CTTTCCAGAGCGCGAGCGTGCACACGACACACCCACCAACGCGCAACCGGTGATAGACT 899
QY 330 uArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAl 350
Db 900 GCGCAACAAACGCTACGAACTCCGCGCGAATGGAAGCAGCCATTCCCGGTTTTTGAAGC 959

```

```

QY 350 aLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlyLeuAlaGlyAspAlaVa 370
Db 960 CCTGGCGGTACATCTGAACCGCAATGACTACCAACAGCGAAGAGCGAGCGATGCGAGT 1019
QY 370 lGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIli 390
Db 1020 AGAAACACTTCTTCAACACAAACACACAAACCGCGTATCGAGTTGCGCCACCAACCCAT 1079
QY 390 eGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSe 410
Db 1080 AGGCCGTCTGAAGGCGAGCTGGGGCTGCAATATTGGGACAAAAATCAGCGCGCTTTC 1139
QY 410 rAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSe 430
Db 1140 CGCCATTCCGAAACCGTCCAAACACCGATGTTGATTGACAAACATGTCCGCCATTACAG 1199
QY 430 rPhePheGlyValGluGlnAlaAsnTyrAspAsnPheThrLeuGluGlyGlyValArgVa 450
Db 1200 CTTTTTCGGGTGTAGAACAGGCAAAATTTGGACAACTTCACGCTTGAAGGCGGCGTACGCGT 1259
QY 450 lGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTy 470
Db 1260 GGAATAACAAAAAGCGCTCCATCCGGTACGACAAAGCATTTGATTGATCGAGAAAACACTCTA 1319
QY 470 rLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSe 490
Db 1320 CAACCGAGCCCTTCGCGCGAGCTCGGGCGCGACCGCAACCGCGCGCTCGTTCGCACTTTC 1379
QY 490 rGlyAsnTyrTyrPheThrProGlnHisLysLysLeuSerLeuThrAlaSerHisGlnGluAr 510
Db 1380 GGGCAACTGGTATTTCACGCCACACCAAACTCAGCCTGAGCGCTCCCATCAGGAACG 1439
QY 510 gLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGl 530
Db 1440 CTTGCGCTCAACGCAAGAACTGTAGCAGACGCGAGCACTGCGCCACCAACACCTTGA 1499
QY 530 uValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTy 550
Db 1500 AGTCGGCAACAAACACCTCAACAAAGAGCGTTCCAAACAATATCGAACTCGCGCTGGGCTA 1559
QY 550 rGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIl 570
Db 1560 CCAAGCGCACCGCTGGCAATACAACTTCGCAACCTTACCGCAACCGATTTCGCAACTACAT 1619
QY 570 eTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMe 590
Db 1620 TTACGCCCAAACTTAAACGACGAGCGCGCGCCCCCAATCCATCGAAGACAGACGCGAAT 1679
QY 590 tLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTy 610
Db 1680 GAAAGCTCGTGGCTACAACCAATCCGTCGCGACTTCTACGCGCGGGAAGCGAAATCTA 1739
QY 610 rPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLe 630
Db 1740 CTTCAAAACGACACCGCGCTACCGCATCGGTGTTCGGCGCATATGTACGAGGCGCGCT 1799
QY 630 uLysAsnLeuProSerLeuProGlyValGluAspAlaTyrGlyAsnArgProPheIleAl 650
Db 1800 GAAAAACCTGCGCTCCCTTACCGCGGAGGAAAGTCCCTACGCGCAACGTCCTTCAATCGC 1859
QY 650 sGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAl 670
Db 1860 ACAAGCCGACAAAACGCCCCCGCATTCGCGCTCGCGCTTCGCTTCCACTTGAAAC 1919
QY 670 aSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLys 690
Db 1920 CTCGCTAAACCGACCGTATCGATGCCAATTTGAGCTACTACCGCGTGTTCGCGCAAAACAA 1979
QY 690 sLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyValAsnTy 710
Db 1980 ACTCGCGCGCTACGAAACGCGTACCGCGGACCATATGCTCAACCTCGGTGCAAACTA 2039
QY 710 rArgArgAsnThrArgTyrGlyGluTyrTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAs 730

```



```
Db 2040 CCGCGCAATACGGCTATGCGAGTGGAAATTGTACGTCAAAGCCGACCACTGCTCAA 2099
QY 730 nGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 2100 CCAATCGTTTACCCACAGACGCTTCTCTGTATACGCGCAAAATGGCGCGAGCTT 2159
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 2160 TACCGCGCGGTAAACGTAAGTTT 2184

RESULT 13
AAZ53324
ID AAZ53324 standard; DNA; 1020 BP.
XX
XX AAZ53324;
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 149 partial DNA sequence SEQ ID NO:597.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
XX Neisseria meningitidis.
OS
XX
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX
XX 31-JUL-1998; 98US-0094859.
XX
XX 02-SEP-1998; 98US-0098994.
XX
XX 02-SEP-1998; 98US-0099062.
XX
XX 09-OCT-1998; 98US-0103749.
XX
XX 09-OCT-1998; 98US-0103794.
XX
XX 09-OCT-1998; 98US-0103796.
XX
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
XX
XX P-PSDB; AAY74562.
XX
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX
XX Claim 7; Page 418-419; 1453pp; English.
XX
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 1020 BP; 274 A; 344 C; 226 G; 176 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 7,94e-153 Length: 1020
Score: 1808.00 Matches: 336
Percent Similarity: 99.41% Conservative: 1
Best Local Similarity: 99.12% Mismatches: 2
Query Match: 44.80% Indels: 0
DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53324 (1-1020)

QY 420 MetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp 439
Db 1 ATGCTGCTTGACAAACAAGTGCACATTTACAGCTTTTTCGGTGTAGAACAGCAAACTGG 60
QY 440 AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyr 459
Db 61 GACAACTTCAGCTTGAGAGGGGTACGGGTGGAANAACAAAGAGCTTCATTTCAGTAC 120
QY 460 AspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyVala 479
Db 121 GACAAAGCATTTGATTCGGGAAAACCTACTACAACCCACCCCTGCCCGACCTCGGCGC 180
QY 480 HisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHis 499
Db 181 CACCGCAAAACCGCCGCTCATTCGCACCTTCGGGCAACTGGTATTTTCAGCCCAACAC 240
QY 500 LysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla 519
Db 241 AAACCTCAGCCTGACCGCTCCCATCAGAAACGCTGCGCTCAAGCAAGAGCTGTACGCA 300
QY 520 HisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGlu 539
Db 301 CACGGCAAAACACGCTCGCCACCAACACCTTTGAAGTCGCAACAAACACCTCAACAAAGAG 360
QY 540 ArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeu 559
Db 361 CGTTCCAAACAATATCGAACTCGCGCTGGGTCTAGGAAGCGGACCGCTGGCAATACAACTG 420
QY 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArg 579
Db 421 GCACCTTACCGCAACCGCTTCGGTACTACATTTACGCCCAACCTTTAAACGACGACGC 480
QY 580 GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 481 GGCCCCAAATCCATCGAAGACGACAGCGAAATGAAGCTCGTCGCTACACCAATCCGCGC 540
QY 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle 619
Db 541 GCGGACTTCTACGGCGCGGAAGCGAAATCTACTTCAAAACCGACACCGCGCTACCGCATC 600
QY 620 GlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
Db 601 GCGGTTTCGGCGACTATGTACGAGGCGCTCTGAAAACCTGCTTCCCTACCGCGCAGA 660
QY 640 GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
Db 661 GAAGATGCTCTACGCGCAACCGTCTCTCATCGCAGACGACCAAAATGCCCCCGTGT 720
QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db 721 CCGGCTGCGCGCTTCGGCTTCCACCTGAAAGCTCGCTGACCGCATCGATCGCAAT 780
QY 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db 781 TTGGACTACTACGCGTGTTCGCCCAAAACAACAACTGCCCGCTACGAAACGCGCACGCC 840
QY 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrp 719
Db 841 GGAACCATATGTCAACCTCGGCGCAAACTACCGCGCAATACGCGCTATGCGAGTGG 900
QY 720 AsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPhe 739
Db 901 AATTGTACGTCAAAGCCGACCAACCTGCTCAACCAATCGTTTACGCCCCACAGCAGCTTT 960
```


XX Moraxella catarrhalis BASB107 DNA.
XX BASB107; infection; pneumonia; otitis media; sinusitis;
KW therapy; diagnosis; vaccine; genetic immunization; antibacterial;
KW antibiotic; receptor; ss.
XX
OS Moraxella catarrhalis.
XX
XX WO2000071724-A2.
XX
XX 30-NOV-2000.
XX
XX 18-MAY-2000; 2000WO-EP04618.
XX
XX 24-MAY-1999; 99GB-0012038.
XX 24-MAY-1999; 99GB-0012040.
XX 01-JUN-1999; 99GB-0012674.
XX 01-JUN-1999; 99GB-0012705.
XX 02-JUN-1999; 99GB-0012838.
XX 08-JUN-1999; 99GB-0013354.
XX
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Thonnard J;
XX
XX WPI; 2001-025166/03.
XX P-PSDB; AAB19945.
XX
XX New BASB103-108 polypeptides isolated from Moraxella catarrhalis
PT bacterium, useful for diagnosing and producing vaccines against
PT bacterial infections such as otitis media and pneumonia
XX
XX Claim 8; Page 73-74; 79pp; English.
XX

CC The present sequence is that of Moraxella catarrhalis ATCC 43617
CC DNA encoding BASB107 (see AAB19945), a protein with amino acid
CC sequence homology to Escherichia coli FhuE receptor precursor
CC outer membrane receptor for Fe(III)-coprogen,
CC Fe(III)-ferritoxamine-B and Fe(III)-rhodotorulic acid. The
CC invention provides M. catarrhalis BASB103-108 polypeptides (see
CC AAB19941-46) and polynucleotides (see AAA89212-17), and methods for
CC producing the polypeptides by recombinant methods. Claimed vaccine
CC compositions comprise a BASB103-108 polypeptide, or a BASB103-108
CC polynucleotide for genetic immunization. A claimed method for
CC diagnosing a M. catarrhalis infection involves identifying a
CC BASB103-108 polypeptide, or an antibody immunospecific to a
CC BASB103-108 polypeptide, in a biological sample. Compositions
CC comprising an immunologically effective amount of a BASB103-108
CC polypeptide, or a polynucleotide encoding it, are used to generate
CC an immune response in an animal. An antibody directed against a
CC BASB103-108 polypeptide can be used to treat humans with M.
CC catarrhalis disease. The polynucleotides may also be used to
CC generate probes and primers, as research reagents for the
CC discovery of therapeutics and diagnostics such as antibacterial
CC agents, and in the design of antisense sequences.
XX

SQ Sequence 2457 BP; 744 A; 560 C; 554 G; 599 T; 0 other;

Alignment Scores:
Pred. No.: 2,87e-132 Length: 2457
Score: 1585.00 Matches: 338
Percent Similarity: 58.92% Conservative: 141
Best Local Similarity: 41.57% Mismatches: 272
Query Match: 39.27% Indels: 63
DB: 22 Gaps: 13

US-09-936-377-2 (1-758) x AAA89216 (1-2457)

QY 7 LysProIleValLeuSerIleIeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
DB 22 AAACCGCTTGCTGTGCCATATTGGCAACTTTTTCATGTCCCAATGCTGGCAGCGCAAT 81

QY 26 ---GluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
DB 82 TTAAGGATAAGCAACCGTCACTTTAGATGGCTTCGATCACCTCTTTTGTACCA 141
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
DB 142 AATACAGATTGGGTAAATCAATCAAAACAGTCAGTCAGTCATCACAGTTTCAAAGAG 201
QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
DB 202 CAACCTACAACCAACGACCAACCTAGCGATGCTTGGCAGGTGAGCTTGGCGTTTCAT 261
QY 85 AlaSerGlnTyrGlyGlyValAspAlaProValIleArgGlnThrGlyArgArg 104
DB 262 TCTAACCATTTTGGGGCGGTGCTCAGCCCATCATTCGTGGCAGAGGGTAAACGC 321
QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
DB 322 CTGAAATCTTACAAACCGGTTTCAGAGGTTCTGGACATGTCTGGGTTCGCGCAGACCAT 381
QY 125 AlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGlyProValThr 144
DB 382 GCCATAGCGGTGGACACCACTGGCAAAACAGGTGGAGATTGCGAGGCTCTGCTGCC 441
QY 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
DB 442 TTGTGTACGCTCTGGCAACTCAGCGGTGGTCAATGTGTTGATGACAAAATACCC 501
QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
DB 502 AGCAATTTGCCAGCAAA---TTACAGGTGATGTGACGGTGGCTTTAGCAGTCCCAAC 558
QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAspPheValLeuHis 204
DB 559 CGTGAATAATTAATCAGCGCAGTGGCAGACCCCACTGGGAGAGCATGTGCGAGTGGCT 618
QY 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
DB 619 GTTGCAGGCTGTCCAAACAGCAGCAGACTATAAACGCCACGCTTTTGACGCGCATGTC 678
QY 221 -----ArgAsnLeu 223
DB 679 TTTAACAAAAACATGAGATGATAACACTCAGCCAGCAATTCATCTATAAGACACCTTA 738
QY 224 LysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
DB 739 AAGCATCTGCCAGACAGCATGCCAAATCAAAACGCGAGCAACGCT-TGGCGTGTCTAGGT 797
QY 244 TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPr 263
DB 798 TGGCAATCAAGCTTTTGGGGGCGATCGTGAGCTTACGCCGAGACAATATGGCGCTGCC 857
QY 263 oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuI 283
DB 858 CAACCATTCATGATATGAAGATGTAGCGTGCATGGCAATTTCTCAGTCGCGCTTACA 917
QY 283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
DB 918 ATACAAGCCATATTGGCTTTGTATCTCTTTTGTATGAAAAATGATGATGATGATGAT 977
QY 303 PAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHis 318
DB 978 CAATGCGGGCTTGAATGCCATACATGATGACCAACGACCAACGACGACGACGACGAC 1037
QY 318 sAlaHisAlaHis-----AsnGlyLysProTrpIleAspLeuArgAsnLys 333
DB 1038 TGACCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1097
QY 333 sArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgVa 353
DB 1098 GCGTTATGATGCAAGGCGCAATCAATCGCCGCTTGTCTGCGCATGTGATAAATCCGAGC 1157
QY 353 lHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPh 373

```

Db 1158 CAGCATGGGCAAGTGGATTATCATCATGATGAGATAGATGGGGGTGAGAGACCAAGCTT 1217
QY ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPhe 390
Db 1218 TTTTGATAATCAAGCAATGTGGCGTCTGGAAGCCTCACATACCCCATTCATAGCC 1277
QY 391 ----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
Db 1278 GATGGGCAAGTTTACGGGGGTGTTTGGGGTGGTATCTCACCTCAAAAACAGCGGACT 1337
QY 409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
Db 1338 TGTGCCACCTCGTTATGACATGCGATTAACAAGACACCCCAAAACATCTTGACAAATAA 1397
QY 424 nLysValGlnHisTyrSerPheGlyValGluGlnAlaAsn---TrpAspAsnPheTh 443
Db 1398 TAAACCAAAACAGGCAGTGTGTTTGGTTTGAAGAATAACAACCCCAATGATCAAGCTGAC 1457
QY 443 rLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
Db 1458 CGTTGACGGCGCGCTCGCATTGAGAAACAACCATCATCGATGATTATGATAAAGCGC 1517
QY 463 uille-----AspArgGluAsnTyrTyrIly 471
Db 1518 CATTTATCAGAGCTTAACTTAGCTTAGCAACGGCTCATGAACACAGACATACGCTTTAA 1577
QY 471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
Db 1578 ACGATTGCTGACAGCGGTACTTTAAACCCCAAAACAAACCGCACGCTCTTATGCTGT 1637
QY 489 uSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
Db 1638 TGGGAGCATTTACAATTACGCCCAACAATAAATTATCGCTGAATCTGTCGATCAAGA 1697
QY 509 uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
Db 1698 ACGCTGCCAAATGCTCAGGAATGTATGCTACGCGCATGCCCTTGCCACCAACCTCGTT 1757
QY 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
Db 1758 TGAATTTGGCAACCGCTTTTAAACAAAGAAATAATCAACATTCATTTGGGGCTGAC 1817
QY 549 YTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
Db 1818 ATTTCAAGGTGATAAATGGATTATCGTCTGGGGCTATCATATGATTTTGATAACTA 1877
QY 569 rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588
Db 1878 TGTGTTTTTACAAACATTTGCGCAGTATAGCAAGGTTTGGCGGCATGCGTCATGATAA 1937
QY 588 rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGl 608
Db 1938 AGACTTAAACACCGCAGCTATGAACAGACGCGCGGAAATTTATGATTTGATGTCAA 1997
QY 608 uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl 628
Db 1998 CATCGGTATCAGATTAAATGATGATATCATGTGGCGGTATTTGGTGATATATTCGTGG 2057
QY 628 YArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
Db 2058 CAAGCTCACCAATTTCCCTGACAAAAAGGCGAGAACCGATGCGTATGGCAACCGTCTCT 2117
QY 648 eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db 2118 CATCAACAGCCAGCAGTCATACGCCAAGACTGCCCAACCAACAAACGCTTGGCATGAAAT 2177
QY 668 uLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGl 688
Db 2178 AACCGCAATGTTAATGCAATTTGGTGGGTTTTTGGATATTCGCCATACCTTTAAACA 2237
QY 688 nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708

```

```

Db 2238 AGATAAATGGCGAATTTTGAACGCCCAACCCAGCTCATAACTTGGTGAATTTGGGGCT 2297
QY 708 aAsnTyrArgArgAsnThrArgTyr-----GlyGluTyrAsnTyrTyrValLysAlaAs 726
Db 2298 TAACTATCAGCACCAAGCCAGCCATCAAGCAGGCTCGGTTCAAGTATTTTTTAATGCTAA 2357
QY 726 pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db 2358 CAATCTATTAAACGATAAAGTCTTGTCTCATGAGACATTTTCCAGACATGCCACAAAT 2417
QY 746 tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2418 GGGCGGCAAACTTATGCTCGGGGCAAAATTTCAAATTT 2454

RESULT 17
AAL46518
ID AAL46518 standard; DNA; 2646 BP.
XX
AC AAL46518;
XX
DT 19-JUL-2002 (first entry)
XX
DE M catarrhalis MCA102062 gene SEQ ID NO: 51.
XX
KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;
KW auditory; antibacterial; otitis media; sinusitis; pneumonia; gene; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200218595-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-CA01221.
XX
PR 28-AUG-2000; 2000US-228294P.
PR 28-AUG-2000; 2000US-228295P.
PR 28-AUG-2000; 2000US-228296P.
PR 29-AUG-2000; 2000US-228438P.
PR 29-AUG-2000; 2000US-228439P.
PR 29-AUG-2000; 2000US-228440P.
PR 29-AUG-2000; 2000US-228441P.
PR 29-AUG-2000; 2000US-228442P.
PR 29-AUG-2000; 2000US-228443P.
PR 29-AUG-2000; 2000US-228511P.
PR 29-AUG-2000; 2000US-228512P.
PR 29-AUG-2000; 2000US-228742P.
PR 29-AUG-2000; 2000US-228773P.
PR 01-SEP-2000; 2000US-229465P.
PR 01-SEP-2000; 2000US-229474P.
PR 01-SEP-2000; 2000US-229475P.
PR 01-SEP-2000; 2000US-229478P.
PR 05-SEP-2000; 2000US-229740P.
PR 05-SEP-2000; 2000US-229803P.
PR 05-SEP-2000; 2000US-229804P.
PR 05-SEP-2000; 2000US-229805P.
PR 05-SEP-2000; 2000US-229806P.
PR 05-SEP-2000; 2000US-229809P.
PR 05-SEP-2000; 2000US-229811P.
PR 06-SEP-2000; 2000US-230214P.
PR 06-SEP-2000; 2000US-230250P.
PR 06-SEP-2000; 2000US-230252P.
XX
(AVET ) AVENTIS PASTEUR LTD.
XX
PA Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
XX
PI WPI; 2002-401721/43.
XX
DR P-PSDB; AAO17586.
XX
Moraxella polypeptide and polynucleotides useful as vaccine for
PT immunizing a host e.g. humans against disease e.g. otitis media,
PT pneumonia, caused by infection of the bacteria -

```

XX Claim 2; Fig 51; 277pp; English.
 XX The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 2646 BP; 817 A; 583 C; 571 G; 675 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,18e-132 Length: 2646
 Score: 1585.00 Matches: 338
 Percent Similarity: 58.92% Conservative: 141
 Best Local Similarity: 41.57% Mismatches: 272
 Query Match: 39.27% Indels: 63
 DB: 24 Gaps: 13
 US-09-936-377-2 (1-758) x AAL46518 (1-2646)
 QY 7 LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
 DB 110 AAACCGCTTGGCGTTCCTGCAATATGGCAACTTTTCAATGCCAATGCTGGCAGAGCGAAT 169
 QY 26 ---GluThrGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
 DB 170 TTAAGGATAAGCAACCGTCAATTTAGATGGCGTTTCGATCACCCTCTTGGCTGACCAA 229
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
 DB 230 AATACAGAGTTTGGCGTTAATCATTCATCAAAACAGTCAGTGGCATCACAGTTTCAAAGAG 289
 QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
 DB 290 CAACATACACACAGCAACCAACCCCTAGGCGATGCTTGGCAGAGTGTGGCGTTTCAAT 349
 QY 85 AlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
 DB 350 TCTAACCATTTGGGGCGGTGCTTCAGCCCATCATCTTGGGGAGGAGGTAAACGC 409
 QY 105 IleLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 DB 410 CTGAATACTCAAAACGGTTGAGAGTGTGGACATGCTGGGTGTGCGCCAGCCAT 469
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
 DB 470 GCCATAGCGGTGGACACCACTGGCAAAACAGGTGGAGATTGTGCGAGGCTCTGCTGCC 529
 QY 145 LeuLeuThrSerSerGlyValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 DB 530 TTGTTGTAGCCTCTGGCAACTCAGCAGCGGTGGTCAATGCTGTGATGACAAAATACCC 589
 QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
 DB 590 AGCAAAATTTGCCAGCAAA---TTCAAGGTGATGTGACGGTGTGCTTAGCAGTCCCAAC 646
 QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
 DB 647 CGTGAATAATTAATCCCGCAGTCCGAGCCCGACCTGGGAGAGCATGTGCGAGTGGGT 706
 QY 205 ThrGluGlyLeuThrArgLysSerGlyAspThrAlaValProArgTyr----- 220
 DB 707 GTTGACGGGCTGTCCAACAGCAGCAGACTATAAAACGCCACGCTTTGACCGCCATGTC 766
 QY 221 -----ArgAsnLeu 223
 DB 767 TTTAAACAAAAACATGAAGATGATAAACAACCTAGCCAGAAATTCATCTATAAGACACCTTA 826
 QY 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
 DB 827 AAGCATCTGCCAGACAGCCATGCGCAAAATCAAAACGCGAGGAACGCT-TGGCGTGTATGGGT 885

QY 244 TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPr 263
 DB 886 TGCATCATCAAGGCTTTTGGGGCATCGGTGAGCTTACGCCGAGACAAATATGCGCTGCC 945
 QY 263 oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuI 283
 DB 946 CAACCATTCACATGAATATGAAGATGTAGCGTGCATTTCTCAGTCCGCCCTTACA 1005
 QY 283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
 DB 1006 ATACAAAGCCATATTCGGTTTGTATCTTTTGTATGGAATAATGATGACTTAGAGTTGA 1065
 QY 303 pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi 318
 DB 1066 CAATGCGGGCTTGAATGCCATACATGATGACACAGCAGCAGCAGCAGCAGCATGTCA 1125
 QY 318 sAlaHisAlaHis-----AsnGlyLysProTyrIleAspLeuArgAsnLys 333
 DB 1126 TGACCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1185
 QY 333 sArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgVa 353
 DB 1186 CGCTTATGATGTCAAGGCAAAATCAATCAATCGCGCTTTGCTGGCATGTATAAAATCGCAGC 1245
 QY 353 lHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAsnPh 373
 DB 1246 CAGCATGGCAAAATGGATTATCATCATGATGATAGTAGTGGGGTGAAGAGCAGCTT 1305
 QY 373 ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
 DB 1306 TTTTGAATATCAAGCAATGTGTGGCTGTGAAGCCTCAGATACCCCATTCATACGCC 1365
 QY 391 ----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
 DB 1366 GATGGCGAAGTTTAGCGGGTGTGGGTAGGTATCTCACCCTCAAAACAGCGGACT 1425
 QY 409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
 DB 1426 TGTGCCACCTCGTTATGAAGATGGCAATAAACAAGACACCAACAAATCTTGCACAAATA 1485
 QY 424 nLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAspAsnPheTh 443
 DB 1486 TAAACCAAAACAGCGAGTGTGTGTGGTTTGAAGAATAACAAACCAATGCAAGAGCTGAC 1545
 QY 443 rLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
 DB 1546 CGTTCAGCGCGCGCTCGCATTTGAGAAACAAACCATCAGCATGGATTATGATTAAGACGC 1605
 QY 463 uIle-----AspArgGluAsnTyrTyrLys 471
 DB 1606 CATTTATCAGAGCTTAAACTTAGGCTTAGCAACCGCTCATGAACCAAGACATACGCTTAA 1665
 QY 471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
 DB 1666 AGGATGTGTGACAGCGGTACTTTAAACCCCAAAACAAACCAACCGCAGCTCTTATGCTGT 1725
 QY 489 uSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
 DB 1726 TGGGACGCAATTACAAATAACGCCCAACATAAATATCGCTGAATCTCTCCATCAAGA 1785
 QY 509 uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
 DB 1786 ACGCTGCGCAATGTCTAGGAATTGTATGCTCAGCGCATGCACCTTGCACCACTCGTT 1845
 QY 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
 DB 1846 TGAATTTGGCAACCGCTTTTAAACAAAGAAATCCAAACCAACATTTGATTTGGGCTGAC 1905
 QY 549 YTrpGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
 DB 1906 ATTTCAAGGTGATTAATGGGATTATCGTCTTGGGGGCTATCATATGATTTTGTAACTA 1965
 QY 569 rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588

```

Db 1966 TGGTGTTCACAACTGCGCAGTATAGACAGGTTTCGCGGCGATCGTCATATAA 2025
Qy 588 rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyG1 608
Db 2026 AGACTTAAAAACCGCAGCGCTATGAACAGAGCGGCGGCAAAATTTATGGATTTGATGCTCAA 2085
Qy 608 uileTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG1 628
Db 2086 CATCGGTTATCAGATTAATGATTAATCATGTGGCGTTATTTGGTGATTAATTCGTGG 2145
Qy 628 YArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProph 648
Db 2146 CNAAGCTCACCAATTTGCTGACAAAAAGGCGAGAACCGATGCGTATGGCAACCGTCTCT 2205
Qy 648 eileAlaGlnAspAsnGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db 2206 CATCAACAGCGCAGCAGTATGATCGCCAGACTGCCACCAACCGCTTGGCATGAAAT 2265
Qy 668 uLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG1 688
Db 2266 AACCGCCAATGTTAATCAAAATGCTCAGGCTTTTGGATAATCGCCATACCTTTAAACA 2325
Qy 688 nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708
Db 2326 AGATAAATTCGCGAATTTTGAACGCCCAACCCAGCTCATAACTTGGTGAATTTGGGCT 2385
Qy 708 aAsnTyrArgArgAsnThrArgTyr-----GlyGluTyrAsnTyrValLysAlaAs 726
Db 2386 TAACATACAGCACCAAGCCAGCCATCAAGCAGGCTCGTTTCAGGTATTTTAAATGCTAA 2445
Qy 726 pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db 2446 CAATCTATTAACGATAAAGCTTTTGTCTCATGACACATTTTCCAGACATGCCACAAAT 2505
Qy 746 tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2506 GGGCGGAACTTTATGCTCGGGCAAAATTTCAAATTT 2542

RESULT 18
AAF28547
ID AAF28547 standard; DNA; 89047 BP.
XX
XX
AC AAF28547;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #34.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
FN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
XX WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX

```

```

PS
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising a combination of 41 nucleic acid molecules (see
CC AAF28544-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localized infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other;

Alignment Scores:
Pred. No.: 4,06e-130 Length: 89047
Score: 1585.00 Matches: 338
Percent Similarity: 59.92% Conservative: 141
Best Local Similarity: 41.57% Mismatches: 272
Query Match: 39.27% Indels: 63
DB: 22 Gaps: 13

```

```

US-09-936-377-2 (1-758) x AAF28547 (1-89047)
Qy 7 LysProIleValLeuSerIleLeu--LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db 424 AAACCGCTTGCCTGTGCCATATTTGCAACTTTTTCATGCCAATGCTGCAGAGCGAAT 483
Qy 26 ---GluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
Db 484 TTAAGGATTAAGCCACCGCTCATTTTAGATGCGCTTTCATCCTCTTTAGCTGACCA 543
Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db 544 AATACAGAGTTTGGCGTTAATCATTCATCAAAAACAGTCAGTGGCATCACAGTTTCAAAGAG 603
Qy 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db 604 CAACTACAAACAGCAGCAACCCCTAGCGCATGCTTGGCAGGTGAGCTTCGCCCTTCAT 663
Qy 85 AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
Db 664 TCTAACCATTTTGGGGGCGGTGCTTCAGCCCATCATCTTCGTGGCAGAGGTAAACGC 723
Qy 105 IleLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db 724 CTGAAATCTACAAAACCGTTTCAGAGTTGTGGACATGCTCGGTTGTGCGCAGACCAT 783
Qy 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
Db 784 GCCATAGCGTGGACACCACTGGCAAAACAGTGGAGATTGTCGAGGCTCTCGTGCC 843
Qy 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db 844 TTGTTGTACGCTCTGGCAACTCAGCAGCGGTGTCATGTCGTTGATGACAAATATACC 903
Qy 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db 904 AGCAATTTGCCAGCAAAA---TTACAAGGTGATGTAGCGTCTTTAGCAGTGGCAAC 960
Qy 185 LeuGluLysLeuThrSerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db 961 CGTGAAAATTAATCACCGCCAGTCCGAGCCCACTGGGAGAGATGTGGCAGTGGCT 1020
Qy 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db 1021 GTTGACGGCTGTCCAAACAGCAGCAGACTATAAAACGCCACGCTTTGACCGCCATGTC 1080
Qy 221 -----ArgAsnLeu 223
Db 1081 TTTAACAACAAACAGATGATGAACACTAGCCAGCAATTCATCTATATAAGACACCTTA 1140

```

Claim 1; Page 324-345; 545bp; English.

29-AUG-2000; 2000US-228511P.
 29-AUG-2000; 2000US-228512P.
 29-AUG-2000; 2000US-228742P.
 29-AUG-2000; 2000US-228773P.
 01-SEP-2000; 2000US-229465P.
 01-SEP-2000; 2000US-229474P.
 01-SEP-2000; 2000US-229475P.
 01-SEP-2000; 2000US-229478P.
 05-SEP-2000; 2000US-229740P.
 05-SEP-2000; 2000US-229803P.
 05-SEP-2000; 2000US-229804P.
 05-SEP-2000; 2000US-229805P.
 05-SEP-2000; 2000US-229806P.
 05-SEP-2000; 2000US-229809P.
 05-SEP-2000; 2000US-229811P.
 06-SEP-2000; 2000US-230214P.
 06-SEP-2000; 2000US-230250P.
 06-SEP-2000; 2000US-230252P.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 XX
 PI Loomore S, Wang J, Bradley B, Ochs M, Yang Y;
 XX
 XX WPI: 2002-401721/43.
 DR P-PSDB; AAO17585.
 XX
 PT Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX
 PS Claim 2; Fig 49; 277pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 2942 BP; 911 A; 562 C; 632 G; 837 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,53e-98 Length: 2942
 Score: 1205.50 Matches: 290
 Percent Similarity: 46.25% Conservative: 142
 Best Local Similarity: 31.05% Mismatches: 303
 Query Match: 29.87% Indels: 199
 DB: 24 Gaps: 18
 US-09-936-377-2 (1-758) x AAL46517 (1-2942)
 QY 9 IleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlu 28
 DB 83 ATCATACCGAGGTATCATGATAAAACCACTTTGTCGATATCGGCCACTTT 142
 QY 29 GlnSerValcylLeuGluThrValThrValValcylLeuSerArgProArgAlaThrSer 48
 DB 143 GCGATCCAGCGTAGCAGATATACCAAGCTGGTGAGAGCAACCACTTAAG 202
 QY 49 GlyLeuLeuHisThrSer----- 54
 DB 203 GGTGTATTGTAGCTCGCAACCAACCAAAATACAGTTTCTATCTAATGATTCAAA 262
 QY 55 ThrAlaSerAspIleLeuSerGlyAspThrLeuArgGlnIleValAlaValAsnLeuGly 74
 DB 263 CAATCAGTGATCTTACGCTTTTCAAAAGATAAAATTAATATATCGTTCGGCAACTTGGGC 322
 QY 75 AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAla 94
 DB 323 AATGCGTGAAGTGGTGGCTTGGTATTATCATAGTAACCTTTTGGTGGCGGTTTCATCTGCA 382
 QY 95 ProValIleArgGlyGlnThrGluThrGluThrGluThrGluThrGluThrGluThr 114
 DB 383 CCTGTTGCGAGGGAAGAGGGTGTGCGTCTTAAGATTTCACAAATGCAACTGATG 442

115 GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGln 134
 DB 443 ATTGATGTGTCATCAATATCGCTGATCATGTTGTGGCCGACCATACACTTTAGCGTCT 502
 QY 135 GlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGly 154
 DB 503 AAAAGTTGAGCTTGTGCTGCTGATACGCTGTTATATGCTTGGCTGGCATCGCCAGCTGGT 562
 QY 155 LeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly----- 171
 DB 563 GTGATTAATGTTGTTGATGACCTATCCGAATCGTATGCTAGTGTGCTATCCATGAC 622
 QY 172 ---ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer 190
 DB 623 AAAATCGAAGCGAGAGCGATGCTTCGATATAACACAAACACCATGAAAAGCTTGCAACT 682
 QY 191 GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg 210
 DB 683 GCAGGGGTGAGCTTTGGGGTAGGATCGCATTCGCTTCGGGTGGAGGGCTTAAAGCGA 742
 QY 211 LysSerGlyAspTyrAlaValProArgTyr-----ArgAsnLeuLysArgLeuPro 227
 DB 743 GAGGTGATGACTATCAAGTTCCTCCCATTTTCAGGCAGATCGCATGTAGATTATGTCCA 802
 QY 228 AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly----- 243
 DB 803 GGTAGT-----GCAATAACTCTACCGTTGGCATGATTGCGGTCTTATATT 850
 QY 244 TrpArgLysArgPheTyrArgThrTyrSerAspArgAspGlnTyrGlyLeuPro 263
 DB 851 CATGATAATGGGCATATCGGTGCTTCTTATAGCCACCGTAAAGATCGTTATGTTATCCCA 910
 QY 264 AlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIle 283
 DB 911 GGCATATCCACTGCGACGCCAACGAGAGCATTTTATCAATGGCATATATATACAAAA 970
 QY 284 AsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLeuAspValAspTyrAsp 303
 DB 971 TCCCAATTATTTTACCCTTTATCTCTTATGATGGAGGATTCAGATTATGAT--GAC 1027
 QY 304 AsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAsn 323
 DB 1028 AATCTCTCATACGATTCGCGCCACACACAGAGCATATCGGTGAGCATATAATCCCAAG 1087
 QY 324 Gly-----LysProTyrIleAspLeuArgAsnLysArgTyr 335
 DB 1088 GCGTGCCTCATCATGACATCATTCGCTTGGATTGATGATAAAACCAATTCGCTAC 1147
 QY 336 GluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu 355
 DB 1148 GACATTCGTGGCGAGGTATATCGCTTCAAGGTTTGGATAAAATTAAGCTTAAGCTTA 1207
 QY 356 AsnArgAspAspTyrHisHisAspGluLys---AlaGlyAspAla----- 369
 DB 1208 ACTATGCGAGATTATTTATCATGATGAAAAGATGCTGGCAATGAGCAAGCAACCAAT 1267
 QY 370 -----ValGluAsnPhePheAsn 375
 DB 1268 CACAAACCTTCTGAGCGTGATACACGGTGGATAAGGGTCATGCCAGCTCTATTATTTACA 1327
 QY 376 AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly 395
 DB 1328 AAAAAAGGGTAAATGCTGCTTGGAGTTATATCATACACCGACCAACCGCTTATCTGGG 1387
 QY 396 SerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla----- 411
 DB 1388 GTATTGGGTATTGGGTATCAAAACCAAAATCTGCAGCAGGAGCGGTATTGTCCAAGC 1447
 QY 412 -----ThrSerGluAlaValLysGlnPro----- 419
 DB 1448 TATTTTCAATCAGACAGATGGCAAAAGCCCAAGTCAAAACATTAAACCAATATCGT 1507

QY 420 -----MetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGlnAla 437
 Db 1508 CTTTACTTATTAGTTCCTCAATAAATAAAGCCCTTGATTTTGGACCTTGACCACTA 1567
 QY 438 AsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIle 457
 Db 1568 AAGTAATCAATGACTTTTAAAGTGGCGATGGTCTATGAACACAAACACCAAT 1627
 QY 458 ArgTyrAspLysAlaLeuLeuAspArgGluAsnTyrTyr----- 470
 Db 1628 GAATATGACGACGCTTACTTGACCATGCTTTGACGATTTTAAAGTAAAGCAGCTA 1687
 QY 471 LysGlnPro---LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
 Db 1688 AAAGCCTGATCATCTGATTTGACGACATATAAACCAACATGCCCTCTTATGCTGGT 1747
 QY 490 SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
 Db 1748 AGTGCCTTATGGATATTACGCCAATCATCGATTGTCTTGACTCTCATTAACGNA 1807
 QY 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPhe 529
 Db 1808 CGCATTCATCGCGATGGAGCTGTATTATCAAGCGGACATTTGGCAGCAGCTCTTT 1867
 QY 530 GluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
 Db 1868 GAGCATGCAATAAACTTGGTCAAGAAATCGGATAATTATGACTGGGTTTATG 1927
 QY 550 TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
 Db 1928 CATACAGCAGATAAAGTCAGCTATAAAGCAAGCACTTACTATAGCAATTTTGATACTAT 1987
 QY 570 IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGlu 589
 Db 1988 ATTTTAAATGAGAC-----ATTGCCAAAGAGGCAAT 2020
 QY 590 MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIle 609
 Db 2021 TTATACATCAGACGCTAATAACACAGCGGTAAAGTTTATGGTGGAGGTTTCATTA 2080
 QY 610 TyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg 629
 Db 2081 ACTTACCAGCCAAATGCCATCACAGTGTGTGTTTTTTGGTGATATGTCAGAGTAAA 2140
 QY 630 LeuLysAsnLeuProSerLeuProGlyArgGluaspAlaTyrGlyAsnArgProPheIle 649
 Db 2141 ATTGGTGCAATTATCTGATATTAAAGGTAAGCTT---GTATATGCTGGCAGAAAGTGGTT 2197
 QY 650 AlaGlnAspAspGln----- 654
 Db 2198 TATTTGATGATATCAAGGATATGACCGTTGATGATATGTTGATATGATGCGGAT 2257
 QY 654 ----- 654
 Db 2258 GGTGGTTTGACTGTGCTTAAAAACGCTGAACAATGGGACAAATTAACGATAATAAT 2317
 QY 654 ----- 654
 Db 2318 GATTGTAGTACACCAATTAATGTCTATAAAACGGTACAAACCTCAGCGAGGAGGAT 2377
 QY 655 -----AsnAlaProArgValProAlaAlaArgLeuGly 665
 Db 2378 TACGACCGTTTGGCAGCAATCCAACTTACGACCCAGAGTACCGCCAGCGCTTGGGC 2437
 QY 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
 Db 2438 ATTCGTTGGCAAGGATATTTGGTGTGATCTGCTGCAATGCAAGATTTTAACCATGTG 2497
 QY 686 PheAlaGlnAsnLysLeuAla----- 692
 Db 2498 TTTGCACAAAATAAAGTTGCCACCTCAACGGTGGCATTAAACCTCAATTCAAGCAGCCA 2557
 QY 692 ----- 692

Db 2558 GAAGTTGCCAACCCCATGAGAGTCATTGCCGAATCAGCGACTATGGCAGTGATAACAAC 2617
 QY 693 -----ArgTyr-----GluThrArgThrProGlyHisHisMetLeu 704
 Db 2618 CCTTTGATGATGACCCCAAGATATATACAGAAAACAAACGGCAGGATATAATTGCTC 2677
 QY 705 AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLys 724
 Db 2678 AATGTTGGCTTAGATTATACATGTCATATCGTAATGTTGATTATACGCTGTCAATTGGT 2737
 QY 725 AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro 744
 Db 2738 GCGAATAATTTACTTAAATGAACAATCTACATTCAACTCATTTTGGCGTTTGTACCG 2797
 QY 745 GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2798 CAGATGGGCGTAATCTGACITTAGTTAGCGGCTAAATTT 2839
 RESULT 20
 AAF28554
 ID AAF28554 standard; DNA; 269223 BP.
 AC AAF28554;
 XX 04-APR-2001 (first entry)
 DE Genomic fragment #41.
 KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 OS Moraxella catarrhalis.
 XX WO200078968-A2.
 XX 28-DEC-2000.
 XX 16-JUN-2000; 2000WO-US16649.
 XX 18-JUN-1999; 99US-0140121.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Lagace RE, Patterson C, Berg KL;
 WPI; 2001-041427/05.
 Genomic library for identifying diagnostic and therapeutic
 compositions, and for identifying virulence factors, regulatory
 elements and drug targets, comprises Moraxella catarrhalis nucleic
 acids -
 Claim 1; Page 486-545; 545pp; English.
 The present invention relates to a Moraxella catarrhalis genomic library
 comprising of a combination of 41 nucleic acid molecules (see
 AAF28514-AAF28554). The library has a number of uses described in the
 specification e.g. is useful for identifying diagnostic and therapeutic
 compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 aerobic, gram-negative diplococcus, normally found among the bacterial
 flora of human upper airways. M. catarrhalis is known to cause acute,
 localised infections such as otitis media, sinusitis and bronchopulmonary
 infection and life-threatening, systemic diseases including endocarditis
 and meningitis.
 SQ Sequence 269223 BP; 77067 A; 56596 C; 57380 G; 78180 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,3e-95 Length: 269223
 Score: 1205,50 Matches: 290
 Percent Similarity: 46,25% Conservative: 142
 Best Local Similarity: 31,05% Mismatches: 303


```

Db 211709 ATTGGTGCAATTATCTGATATTAAAGGTAAGCTT---GTATATGCTGGCAGAAAGTGGTT 211765
      |||      :|||:      |||      |||      |||      |||      |||      |||
Qy 650 AlaGlnAspGln----- 654
      |||      |||      |||      |||      |||      |||      |||      |||
Db 211766 TATTTGATGATATCAAGGATATGACCGTTGATGATAATGGTGATATGATGCCGAT 211825
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 654 ----- 654
      |||      |||      |||      |||      |||      |||      |||      |||
Db 211826 GGTGGTTTGACTGTGCTTAAACCGCTGAACAAATGGGAGCAAAATTAAACGATAATAAT 211885
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 654 ----- 654
      |||      |||      |||      |||      |||      |||      |||      |||
Db 211886 GATTGTAGTACACCATTAATGCTCTATAAAACGGTACACAACTCAGCGAGGAGGAT 211945
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 655 ----- 655
      |||      |||      |||      |||      |||      |||      |||      |||
Db 211946 TACGACCGTTGGCAGCAATCCAACTTACGACCCAGAGTACGCCCGCGCTGGGC 212005
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212006 AATCGTTGGCAAGGTATTTTGGTGATCATTTGGTCTGCCAATGCCAGAAATTTAACCATGTG 212065
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 686 PheAlaGlnAsnLysLeuAla----- 692
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212066 TTTGCACAAATAAAGTTGGCCACTCAACGGTTGCCATTAACTCAATTCAGCAGCCA 212125
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 692 ----- 692
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212126 GAAGTTGCCAACGCCATGAGAGTCAATGCCGAATCAGCGACTATGCCAGTGATAACAAC 212185
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 693 ----- 693
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212186 CTTTGATGATGACCCAGGATATATCAGAAAAACAAACGGCAGGATATAATTTGCTC 212245
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 705 AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLys 724
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212246 AATGTGGCTTAGATTATAACAATGCATATCGTAATGTTGATTATACGCTGTCATTGCT 212305
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 725 AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro 744
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212306 GCGAATAATTTACTTAATGAACAATCTACATTCACAACCTCATTTTTCGCGTTTGTACCG 212365
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 745 GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
      |||      |||      |||      |||      |||      |||      |||      |||
Db 212366 CAGATGGGGCGTAATCTGACTTTAGTTTTCGCGCTAAATTT 212407
      |||      |||      |||      |||      |||      |||      |||      |||

```

RESULT 21

```

ID AAA89217
XX AAA89217 standard; DNA; 2742 BP.
XX AC AAA89217;
XX DT 19-MAR-2001 (first entry)
XX DE Moraxella catarrhalis BASB108 DNA.
XX KW BASB108; infection; pneumonia; otitis media; sinusitis;
XX KW therapy; diagnosis; vaccine; genetic immunization; antibacterial;
XX KW antibiotic; receptor; ss.
XX OS Moraxella catarrhalis.
XX PN WO200071724-A2.
XX XX 30-NOV-2000.
XX PF 18-MAY-2000; 2000WO-EP04618.
XX PR 24-MAY-1999; 99GB-0012038.
XX PR 24-MAY-1999; 99GB-0012040.
XX PR 01-JUN-1999; 99GB-0012674.
XX PR 01-JUN-1999; 99GB-0012705.

```

```

PR 02-JUN-1999; 99GB-0012838.
PR 08-JUN-1999; 99GB-0013354.
XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Thonard J;
XX WPI; 2001-025166/03.
XX P-PSDB; AAB19946.
XX New BASB103-108 polypeptides isolated from Moraxella catarrhalis
PT bacterium useful for diagnosing and producing vaccines against
PT bacterial infections such as otitis media and pneumonia
XX Claim 8; Page 75-76; 79pp; English.
XX The present sequence is that of Moraxella catarrhalis ATCC 43617
CC DNA encoding BASB107 (see AAB19946), a protein with amino acid
CC sequence homology to Vibrio cholerae haem receptor HcrA. The
CC invention provides M. catarrhalis BASB103-108 polypeptides (see
CC AAB19941-46) and polynucleotides (see AAA89212-17), and methods for
CC producing the polypeptides by recombinant methods. Claimed vaccine
CC compositions comprise a BASB103-108 polypeptide, or a BASB103-108
CC polynucleotide for genetic immunization. A claimed method for
CC diagnosing a M. catarrhalis infection involves identifying a
CC BASB103-108 polypeptide, or an antibody immunospecific to a
CC BASB103-108 polypeptide, in a biological sample. Compositions
CC comprising an immunologically effective amount of a BASB103-108
CC polypeptide, or a polynucleotide encoding it, are used to generate
CC an immune response in an animal. An antibody directed against a
CC BASB103-108 polypeptide can be used to treat humans with M.
CC catarrhalis disease. The polynucleotides may also be used to
CC generate probes and primers, as research reagents for the
CC discovery of therapeutics and diagnostics such as antibacterial
CC agents, and in the design of antisense sequences.
SQ Sequence 2742 BP; 844 A; 536 C; 600 G; 762 T; 0 other;

Alignment Scores:
Pred. No.:      8,47e-98      Length:      2742
Score:          1202.00      Matches:      291
Percent Similarity: 46.66%      Conservative: 142
Best Local Similarity: 31.38%      Mismatches:   301
Query Match:      29.78%      Indels:       194
DB:              22          Gaps:         20

US-09-936-377-2 (1-758) x AAA89217 (1-2742)
Qy 7 LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
      |||      |||      |||      |||      |||      |||      |||      |||
Db 10 AAACCACTTGTGTGCG-----ATATCGGCCACCTTTGCGATGCCAGGTAGCA 60
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 27 ThrGluGlnSerValGlyLeuGluThrValThr-----ValValGlyLysSer 42
      |||      |||      |||      |||      |||      |||      |||      |||
Db 61 GATAATACCAAGCTGGTGAAGAGCAACCAACCACTTAAGGGTGTATTTGTAAGCTCG 120
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 43 ArgProArgAlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAspLysIle 60
      |||      |||      |||      |||      |||      |||      |||      |||
Db 121 CAACGAACCAAAATACAGCTTTTGTATCTAATGATTCAAAACATCCAGTCATCTTAG 180
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
      |||      |||      |||      |||      |||      |||      |||      |||
Db 181 CTTTCAAAAGATAAATAATATGTTTCGGCAACCTTGGGCAATCGCTAAGTGGTGAG 240
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
      |||      |||      |||      |||      |||      |||      |||      |||
Db 241 CTTGGTATTCTAGTAACCTTTTGGTGGCGGTTCACTGACCCTTGTTGTGGAGGGCAA 300
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 101 ThrGlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPhe 120
      |||      |||      |||      |||      |||      |||      |||      |||
Db 301 GAGGTTGGCGTCTTAAGATTTTACAAATGGAACCTGATGATGATGATGATCATCAATA 360
      |||      |||      |||      |||      |||      |||      |||      |||
Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140

```


Qy 693 ArgTyr-----GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyr 710
 Db 2536 AGATATATACAGAAAAACAAACGCGAGATATAATTTGCTCAATGTTGGCTTAGATTAT 2595
 Qy 711 ArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuAsn 730
 Db 2596 AACAAATGATATCGTAATGTTGATTATACGCTGTCAATTCGTGCAATTAATTTACTTAAT 2655
 Qy 731 GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPhe 750
 Db 2656 GAAACAATCTCACTCAACAATCTATTTTGGCGTTGTACCGCAGATGGGCGTAATCTG 2715
 Qy 751 ThrGlyGlyValAsnValLysPhe 758
 Db 2716 ACTTTAGGTTTGACGGCTAAATTT 2739
 RESULT 22
 AAZ53317
 ID AAZ53317 standard; DNA; 522 BP.
 XX AC
 XX AAZ53317;
 XX DT
 XX 21-MAR-2000 (first entry)
 XX DE
 XX Neisseria gonorrhoeae ORF 147 partial DNA sequence SEQ ID NO:583.
 XX KW
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 XX OS
 XX Neisseria gonorrhoeae.
 XX FN
 XX WO9957280-A2.
 XX PD
 XX 11-NOV-1999.
 XX PF
 XX 30-APR-1999; 99WO-US09346.
 XX PR
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX PA
 XX (CHIR) CHIRON CORP.
 XX (GENO-) INST GENOMIC RES.
 XX PI
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 XX P-P8DB; AAY74555.
 XX PT
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX PS
 XX Claim 7; Page 412; 1453pp; English.
 XX CC
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX SQ
 XX Sequence 522 BP; 126 A; 156 C; 135 G; 105 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,97e-63 Length: 522
 Score: 803.00 Matches: 163
 Percent Similarity: 98.81% Conservative: 3
 Best Local Similarity: 97.02% Mismatches: 1
 Query Match: 19.90% Indels: 1
 DB: 21 Gaps: 0
 US-09-936-377-2 (1-758) x AAZ53317 (1-522)
 Qy 1 MetAlaGlnThrThrLeuLysPheValLeuSerIleLeuLeuIleAsnThrProLeu 20
 Db 19 ATGGCACAATCACACTCAAAACCCATTGTTTATCAATCTTTTAAATCACACACCCCTC 78
 Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 Db 79 CTGCCCCAAGCGCATGAACCTGAGCAATCGGTGGGCTTGGAAACGCTCAGCGCTCGTGGC 138
 Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 139 AAAAGCCCTCCGCGCGCGACTTCGGGGCTGTGTCACACTTCGACCGCTCCGACAAATC 198
 Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 199 ATCTCGGCGATACTTTTGGCCAAACCGCTCAACTTGGGCGACGCTTGGACGGCGTA 258
 Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 259 CCGGCGCATCCAGCTTCGCAATACGCGCGCGCGCATCCGCTCCGCTTATTCCGCGTCAA 318
 Qy 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 319 ACGGCGACAGCGATTAAAGTATTGAACCATCACGCGCAACCGGGCGATATGCGGACTTT 378
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 379 TCTCCGATACGCGCATTTATGGTAGATACCGCTTGTGCAACACAGGTTGAATCTCGCG 438
 Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAla-GlyLeuValAspValAla 160
 Db 439 GGGCGGTTAGCTCTTGTACAGCTCGGCAATGTGGCGGCGCTGGTCAATGTTGCCGA 498
 Qy 160 pGlyLysIleProGluLysMet 167
 Db 499 TGGAAAAATCCCCCAAAAAATG 520
 RESULT 23
 ABQ90153
 ID ABQ90153 standard; DNA; 2127 BP.
 XX AC
 XX ABQ90153;
 XX DT
 XX 01-OCT-2002 (first entry)
 XX DE
 XX M. capsulatus gene #138 for DNA array.
 XX KW
 XX Micro array; gene; ds; differential expression; gene expression.
 XX OS
 XX Methylococcus capsulatus.
 XX PN
 XX WO200255655-A2.
 XX PD
 XX 18-JUL-2002.
 XX PF
 XX 14-JAN-2002; 2002WO-NO00019.
 XX PR
 XX 12-JAN-2001; 2001NO-0000235.

QY 522 shisValalaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
 Db 1440 ACACGACGCCACCCAGCTACGAGCTGGGCAATCCCACTCCAGGAAACCTCTTA 1499
 QY 542 nAsnileGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
 Db 1500 CAACCTTCAGCTGGGCTACCGCTACAAGGGCGAGAGTCACGGCGAGCTCAACGTGTT 1559
 QY 562 rArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProly 582
 Db 1560 CCAGAATTTCGTGAACGACTACATCTACCAGCAGATCCAGGATGAG----- 1605
 QY 582 sSerileGluAspAspSerGlu-----Me 590
 Db 1606 -GTGTTCACGAAGATACGGAACAGTTGAGAAAGTGTGCACCAATCCGGGGCGCTGTT 1664
 QY 590 tLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTy 610
 Db 1665 CCCGTGTGTGACACCACTCAGCGCGGCGGTATTCTCTGGCTACGAAGCAACGTGAC 1724
 QY 610 rPheLysProThrProArgTyrArgIleGlyVal-----SerGlyAspTy 625
 Db 1725 ATTT---CCGTGTGTGAGAACCAATTACGCGCTGATCGACCTGCACCTGTTTCAGCGACT 1781
 QY 625 rValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAs 645
 Db 1782 CACCGCGCGCGCTGCTCAAC----- 1803
 QY 645 nArgProPheIleAlaGlnAspGlnAsnAlaProArgValProAlaAlaArgLeuG 665
 Db 1804 -----GGCGGCAACGTGCGGCGCATGCGCGCTGGCTACGG 1841
 QY 665 yPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVa 685
 Db 1842 CTTCAGCTGACTATGCTGCGAGCAGCATGAGCGGCAATCTCGGCTGACTCGCGG 1901
 QY 685 lPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAs 705
 Db 1902 CGAGAGCAGATCTATCGCGGCGAGACGACGCCATGACCAACGGCTAGCTGCTCTCAA 1961
 QY 705 nLeuGlyAlaAsnTyrArg---ArgAsnThrArgTyrGlyGluTrpAsnTrpTyrVally 724
 Db 1962 CTTGGTGTGCAATACGAGTCAAGGCATTCAAGGAGCGCGAAGTCTGCTTCTGCTCA 2021
 QY 724 sAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSer---AspTh 743
 Db 2022 AGGCAACAACCTGCTCAACAGACACCTCCGAAACTCCACCTCTATCTCGCTATTTCGC 2081
 QY 743 rProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2082 GCCGAGGCGGCGCGGCTGAGTTGGGCTACGATCACTAT 2127
 RESULT 24
 ID AAF26431
 AC AAF26431 standard; DNA; 1173 BP.
 AC AAF26431;
 XX 02-MAY-2001 (first entry)
 DT Pseudomonas sp heavy metal transporter encoding DNA ORF08372.
 DE Heavy metal transporter; iron transporter; transgenic plant;
 KW homeostasis regulator; heavy metal ion; trace element; soil pollution;
 KW plant growth promoter; plant development; ds.
 OS Pseudomonas sp.
 XX DE19934720-A1.
 PN 25-JAN-2001.
 XX 23-JUL-1999; 99DE-1034720.

XX 23-JUL-1999; 99DE-1034720.
 XX (TIGR-) TIGR INST GENOMIC RES.
 PA (QUTA-) QUITAGEN GMBH.
 PA (GBFE) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX WPI; 2001-160596/17.
 XX New DNA encoding iron or heavy metal transporters, useful for promoting
 PT growth of plants on polluted soil -
 PS Claim 3a; Page 21-22; 54pp; German.
 XX This invention describes novel DNA sequences (I) that encode iron or
 CC heavy metal transporters (II). The invention also describes (1) a
 CC recombinant expression vector (III) containing (I); (2) prokaryotic or
 CC eukaryotic cells (III) transformed with (I) or (III); (3) production of
 CC (II) by culturing (III); (4) (partial) expression products (IV) of (I)
 CC and synthetic proteins or peptides with the same sequences; (5)
 CC antibodies (Ab) specific for (IV); (6) hybridoma cells that produce
 CC monoclonal Ab; and (7) transgenic plants that contain (III). The iron or
 CC heavy metal transporters encoded by (I) transport heavy metal ions across
 CC cell walls and regulate homeostasis of trace elements. (I), and their
 CC fragments are useful for: (1) expression of (II); (2) as probes and
 CC primers for detection, isolation and amplification of full length cDNA
 CC sequences; and (3) producing transgenic plants. (III) are used to promote
 CC growth, development and yield of plants, particularly leguminosae, to
 CC especially when growing in soil polluted by heavy metals injurious, to
 CC plants. They also improve homeostasis of iron and trace elements. Host
 CC cells that express the iron or heavy metal transporters take up heavy
 CC metals, so reduce pollution of soil and release iron and other trace
 CC metals which improves soil quality and protect plants against pollutants
 CC from the soil.
 XX Sequence 1173 BP; 236 A; 378 C; 362 G; 196 T; 1 other;
 SQ Alignment Scores:
 Pred. No.: 1-01e-14 Length: 1173
 Score: 271.50 Matches: 96
 Percent Similarity: 42.15% Conservatives: 49
 Best local Similarity: 27.91% Mismatches: 148
 Query Match: 6.73% Indels: 51
 DB: 22 Gaps: 12
 US-09-936-377-2 (1-758) x AAF26431 (1-1173)

QY 430 SerPheGlyValGluGlnAlaAsnTrp-----AspAsnPheThrLeuGluGly 446
 Db 85 GCGCTGTTTCATGCTGGAGCAG-----TGCGAAGCCACCGACCGCTGAACCTGAGCCTG 138
 QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArg 466
 Db 139 GCGCGCGCATGAG-----CACCCCGCTAGACCCGACCGCCAAAGGCAAC 196
 QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSer 486
 Db 187 GAGAACTTCGTCGAC-----GGCAGACGCGCCAGCTTCAACGCC 228
 QY 487 PheAlaLeuSerGlyAsn-----TrpTyrPheThrProGlnHisLysLeuSerLeuThr 504
 Db 229 TTCAGCTGTGTCGCGGCGCTGTGTACCAAGCTCGACCAATCTGTCGCTGCGCGCAAC 288
 QY 505 AlaSerHisGlnArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisVal 524
 Db 289 CTCGCTACACGACGCGCGCCCTTACGAGCTGTACGCCAATGCGCGCACGTA 348
 QY 525 AlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnIle 544
 Db 349 GCCACCGCGCGCTTGAAGTCGCTGACGCCACCTGACACAGGAAAGGCCATTTCGCC 408

QY 184 AsnLeuGluLys-----LeuThrSer 190
D 292200 TTATCGAAAAGGATGTTCTGTAATTTGGCGAAATGCAAAATTCGATGTTCTTATTAGT 292259
QY 191 Gly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGlu 206
D 292260 GGTTCCTTAATAATCGGATATATTTACGCAGTGGTAAGGCNACAAAGCTTAATAATATACC 292319
QY 207 GlyLeuTyArgLysSerGlyAspTyrAlaValProArgTyrArg-----Asn 222
D 292320 GCC---TATAACAGTTTGGGGCTTAGCAAAATTCGGTTGGCAAAATTAATGATGCGAAC 292376
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
D 292377 CGCGTGGAAATTATCCACCGCAAACTCGTTTAA--ACAACAAGCAGCACAAGCAATAATG 292434
QY 243 GlyTyrArgLys-----ArgPheTyrArgArgTyrTyrSerAsp 255
D 292435 AGGTGAAAACGAACTTACTAATGAACAAATTAACAGATCAAAATCAAAAGTTCACCGAC 292494
QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAsp----- 271
D 292495 AAAAAAGCATCTCTCTCTCTCTACAAACACCAACCATCACATCAGAAAGATCAGAGTTT 292554
QY 272 -----CysHisAlaAspIleIle-----TrpGlnLysSerLeu 282
D 292555 ACTCTAAAGTGAACACAGCTTTAGTAGTGTCAGTTATTATTAAGTGCATCAAAATTCCTG 292614
QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuTyrGluGluAspValAspTyr 302
D 292615 ATCAAGCAGCGTATTTAAC-TATTAT-----TTAAGCGCA----- 292649
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHis 322
D 292650 GATAATCCTTATCTA-----AATACGCATATCGCACTGAT 292685
QY 323 AsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
D 292686 AACATTAACATTTAGAAAGAACAGCGTAAGTCGTGCTGTAAGATCAAGATCACTAAA 292745
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
D 292746 TTGACTACCGAGGT-----ATAAATTACGTAATCTTCGGAATTATCT 292790
QY 359 -----AspTyrHisHisAspGlyLysAlaGlyAspAlaVal 370
D 292791 CACATTTCTCTTTTATGGGGTGGATATATATCGAGATAAAATCCGTACCGAA----- 292844
QY 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle 390
D 292845 -----CGAGGCACAAACGGTAGCGATGCGAAGTTTCAGCGGACCCCTAT 292889
QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerAlaLeuSer 410
D 292890 AATGCGAATTCAAACACTACAGCGGTT---TATTTA----- 292922
QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
D 292923 -----ATCGCCCATATTCCG 292937
QY 431 PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgVal 450
D 292938 CTATTTGGG----- 292946
QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
D 292947 GAAAAATTTGCTAGTTTCGCAAGTAGTACGTTATGACCACCTACGATCACTCAAGTAAACC 293006
QY 463 LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
D 293007 GTAAAAATCAAGGATAATCATTTA----- 293030

QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSer 502
D 293031 -----TCCTCTCCCAAAATAAATTAATCTGGATAGTACC-----AATTGGTATAGAT 293075
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
D 293076 TTTACTGCGCAATATATTAATGAAGCTTTCCGAGCACCATCTATGCAAGAGCGATTTGTGAGT 293135
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
D 293136 GGTGCTCACTTTGGGCAAAATACTCTAGGCTAGATCACATCAATAGATTTGTAGCAAAAT 293195
QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
D 293196 CCAAAATTTGGCCCTGAAACAGCGAAAAATAAAGAAATACCCGCAAAATCTACATTTTGTAT 293255
QY 547 AlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
D 293256 AGTCTCTTTAAACAAGCGGATAAATTCAAAATTTGAAGCGACTTATTTCGTAATGATGTG 293315
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
D 293316 AAAGATTTTATTAACTTAAATAATTTAATGATGCAAGACAAGTGCAGTGCAGGTGCA 293375
QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595
D 293376 AATCCAAATACAAATAGGAGATTTGTCGCAAAATTTCCAGATCAAA----- 293423
QY 596 AsnGlnSerGlyValAlaAspPheTyrGlyAlaGluGlyIleTyrPheLysProThrPro 615
D 293424 AACATTAACATTAATGCGGTTTAAAGCGTATTGAATTCGAAGTCAATACCAA---ACAGAA 293480
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
D 293481 CGTTTAAACGCTA-----TTTACTAATATGGCAGC 293510
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
D 293511 ACCAAAGGTAAGATAAGATAGTGGC-----GAAAGCT 293543
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
D 293544 TTATCAACATTTCCCGCAGCAAAATCGCGTAGGGTAAATTTATGCTTTAGTAAAGAC 293603
QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
D 293604 AAATTCACGTTGGAGCGACAGTAACCCATTACGCT-----GCTCAACGCGAGTGCCT 293657
QY 693 ArgTyrGluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
D 293658 AAAGATCATAGTGTATACCTACCCCAAGTTATATATACCTGACCGATCTTCGTCTACCTATGCA 293717
QY 712 ArgAsnThrArgTyrGlyGluTyr 719
D 293718 -----CCATTAAAGCGGAATGG 293735
RESULT 26
AAS97293
ID AAS97293 standard; DNA; 2145 BP.
XX
AC AAS97293;
XX
DT 12-MAR-2002 (first entry)
XX
DE Neisseria meningitidis virulence gene #98.
XX
KW Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
XX infection; Gram-negative bacteria; antimicrobial; ds.
OS Neisseria meningitidis.
XX
FN WO200185772-A2.
XX


```

QY 460 pLysalaleuileAaspArgGluAsnTyTyTyLysGlnProLeuProAaspLeuGlyAlaHi 480
Db 1332 CGGCAAAACCGTTTCAAGCAACACCTTAAC-----CCAGTTTCGGCGTGAT 1379
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyPheThrProGlnHisly 500
Db 1380 TTGGCAGCGCACGAA-----CACTGGAGCTTCAGCGCAGCCACAA 1421
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyAlaHi 520
Db 1422 CTAC-----GCCAGCGCAGCGCGCGCTG-----TATGAGCGCTGCAAAACCA 1466
QY 520 sGlyLyshis-----ValAlaThrAsnThrPheGluValGlyAsnLyshis 536
Db 1467 CGCACAACCGGCATCATCTCGATTGCCAGCGCAG----- 1503
QY 536 uAsnLySgluArgSerAsnAsnIleGluLeuAlaLeuGlyTyTyGluGlyAspArgTrpGl 556
Db 1504 -AAAGCGGAACGCGCGCAATACCGAAATCGCTTCAACTCAACGACGCGCAGCTTGC 1562
QY 556 nTyraAn-----LeuAlaLeuTyTyArgAsnAr 565
Db 1563 CGCAACGCGCAGCTACTTCTGCGACGACCATCAAGACGCGCTTGCCCAATCCGCAAAACG 1622
QY 565 gPheGlyAsnTyTyIleTyTyAlaGlnThrLeuAsnAspGly----- 578
Db 1623 CCACGACTCTGCGCGTTCGGTGAAGCGTCAATGCCGTTACATCAAAACACGCGTTA 1682
QY 579 -----ArgGl 580
Db 1683 CGAATTGGCGCGTCTCTACCGCAGCGCGCTGCTGCTGCAAGTCGGGTGAAGCCACAG 1742
QY 580 yProLySerIleGluAspSerGluMetLySleuValArgTyTyAsnGlnSerGlyAl 600
Db 1743 CAACCGCGCTTTTACGATACGACCAAGCAAGCTGTTGAGCGCGAAATCTTGAA----- 1797
QY 600 aAspPheTyTyGlyAlaGlu-----GlyGluIleTyTyPhelysPr 613
Db 1798 -----TTTGGGCGACAGTGGCGGCTTGGACGGTTCCTTGCCTACGCTTCCA 1850
QY 613 oThrProArgTyTyArgIleGlyValSerGlyAspTyTyValArgGlyArgLeuLyAsnLe 633
Db 1851 AAACCCGAATCTGGAATCGGTCGGCGCGCGCTATGTT----- 1890
QY 633 uProSerLeuProGlyArgGluAspAlaTyTyGlyAsnArgPropheIleAlaGlnAspAs 653
Db 1891 -----CAAAAGCCGCTGGTTCGATATTGGTGGCAGGTCAAAAGA 1931
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLySAlaSerLeuTh 673
Db 1932 CGGCAACGGCAATTTGAAAAGCTTGACGAAAGTTTCGGTGTGAAC----- 1980
QY 673 rAspArgIleAspAlaAsnLeuAspTyTyArgValPheAlaGlnAsnLyLeuAlaAr 693
Db 1981 -----GATGTCTTCGCCCACTGGAAACCGCTGGG 2009
QY 693 gTyTyGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyTyArgAs 713
Db 2010 CAAGACACGCTCAAT-----GTTAATCTTCGGTTAAC----- 2043
QY 713 nThrArgTyTyGlyGluTrpAsnTrpTyTyValLySAlaAspAsnLeuLeuAsnGlnSerVa 733
Db 2044 -----AACGTGTTCACACGTTCTA 2063
QY 733 lTyTyAlaHisSerSerPheLeuSerAspThr-----ProGlnMetGlyArgSerPheThrGl 752
Db 2064 CTATCGCACACCAACGATGACCAATACCTCGCGCGCGTGGGACGTGATGATGCGCTT 2123
QY 752 yGlyValasnValLySphe 758
Db 2124 GGGCGTGAACCTCAAGTTT

```

```

RESULT 27
AAA81477/c
ID AAA81477 standard; DNA; 26778 BP.
XX
AC AAA81477;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_25 SEQ ID NO:25.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 524-531; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81360 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 26778 BP; 6602 A; 7008 C; 6846 G; 6318 T; 4 other;

```

```

Alignment Scores:
Pred. No.: 4.48e-09 Length: 26778
Score: 229.50 Matches: 197
Percent Similarity: 33.45% Conservative: 93
Best Local Similarity: 22.72% Mismatches: 301
Query Match: 5.69% Indels: 277

```


Db 4628 -----TTTGGCGCACAAAGTCGGCGCCTTGGACCTTCCCTTGGCTACCGCTTCCA 4576
 Qy 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLe 633
 Db 4575 AAACCCGAATCGGAATCGGCTGCGCGCGCGTATATGTT----- 4536
 Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
 Db 4335 -----CHAAAGCCGTGGTTCGATATTGGTGGCAGGTCAAAAGA 4495
 Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
 Db 4494 CCGCAACCGCAAAATGGAAAACGTTGTACGCAAGGTTTCGGTGTGAAC----- 4446
 Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
 Db 4445 -----GATGTCTTCGCCCACTGGAAACCGCTGGG 4417
 Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
 Db 4416 CAAGACACGCTCAAT-----GTTAACTTTTCGGTTAAC----- 4383
 Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerVa 733
 Db 4382 -----AACGRTTCAACACGCTTCTA 4363
 Qy 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
 Db 4362 CTATCCGCACAGCCACACGATGACCAATACCCCTGCCGGCGTGGACGTGATGACGCTT 4303
 Qy 752 yGlyValAsnValLysPhe 758
 Db 4302 GGGCGTGAACCTACAAGTTC 4284

RESULT 28

AAF21612
 ID AAF21612 standard; DNA; 349980 BP.
 XX
 AC AAF21612;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200066791-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05928.
 XX
 PR 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX
 DR WPI; 2000-647603/62.
 XX
 PT Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -
 XX
 PS Claim 7; Appendix A; 692pp; English.

XX
 CC The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisseria bacteria or as a diagnostic reagent for detecting the
 CC presence of Neisseria bacteria or of antibodies raised to Neisseria
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.
 XX

SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Alignment Scores:

Pred. No.:	1.55e-07	Length:	349980
Score:	229.50	Matches:	197
Percent Similarity:	33.45%	Conservative:	93
Best Local Similarity:	22.72%	Mismatches:	301
Query Match:	5.69%	Indels:	277
DB:	21	Gaps:	46

US-09-936-377-2 (1-758) x AAF21612 (1-349980)

Qy	16	IleAsnThrProLeu	-----LeuAla 22
Db	292611	ATGAATACCCCAATGTTCCGCTCTCAGCTCTGCTTACCTTGGCGGAGGTTTGGC	292670
Qy	23	GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer	42
Db	292671	CATCGGCAGAAATAATGCAAGTGTCTGATACCGTACCGTAAAGGCGACCGC	292730
Qy	43	ArgProArgAlaThrSerSerGlyLeuHisThrThrAlaSerAspLysIleIleSer	62
Db	292731	CAA-----GGCAGCAAAATCCGTACCAACATGTT--- 292760	
Qy	63	GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp	78
Db	292761	-----ACGCTGCAACAAAGAGCAGCAACCGCATATGCGGAACTCTTAAAA	292814
Qy	79	GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIle---	97
Db	292815	GAAGAGCCCTCCATC-----GATTCGGCGCGCAACGCGACGCTCCCAATTCCTG	292865
Qy	98	-----ArgGlyGlnThrGlyArgArgIleLysVal----- 107	
Db	292866	ACGCTGCGCGCATGGGTCAAACTCTGACATCAAGTGGACACGCTATTCCGAC	292925
Qy	108	-----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis	124
Db	292926	AGCCAAATCTTTTACCACCAAGGAGGAG----- 292952	
Qy	125	AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr	144
Db	292953	---TTTATTGTCGATCCCGCTTTGGTTAAAGTGTGTTCCGTACAAAAGGC-----	293000
Qy	145	LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro	164
Db	293001	-----GCGGTTCCGCTCTGCGGTTATCGCGGACCAACGCGCGGATCATC	293048
Qy	165	GluLysMetProGlu-----AsnGlyValSerGlyGluLeuGlyLeuArg	179


```

Db 293049 ACCAAACCGCTGATGCGCCACAGACCTGCTCAAGGCTTGGATATAAAACCTGGGGCGCTGGCG 293108
Qy 180 LeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyLeuAsnIleGlyLeu 197
Db 293109 CTCACAGCGGCTTGGCCAGCAACGAGCGGTAAAGCTAGCGGCGCAAGCTATTCGGGAAA 293168
Qy 198 GlyLysAsnPheValLeuHisThrGluGlyLeuThr-----ArgLysSerGly 213
Db 293169 GAGGCAACCTTC-----GACGGCTTGTCTCTTACACCGCAACAATGAAAAA 293216
Qy 214 AspTyrAlaValProArg---TyrArgAsnLeu-----LysArgLeuProasp 228
Db 293217 GATTACGAGCGGTAAAGCTTCGTTAATAATTCAACGGCGGCAAAACCGTACCGGTAC 293276
Qy 229 SerPro-----ArgArgPheAlaAsnGlyGln 237
Db 293277 AGCGCGCTGGCAACCGCAGCTACCTCGCCAAAATCGGAACAAGCTTCGGCGACGCGAC 293336
Qy 238 HisArgAlaValLeuGlyTrpArgLysArgPheTyrArgThrTyr-SerAspArgArg 257
Db 293337 CACCGCATCGTATTG-----AGCCATATGAAA 293363
Qy 258 AspGlnTyr---GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIle 276
Db 293364 GACCAGCACCGGGCATCGGTACCGTCCGTGAAGATTT-----ACCGTC 293408
Qy 277 IleTrpGlnLysSerLeuIleAsn---LysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
Db 293409 GCGCGCGATAAGAGCGAATAAGTATGGAACCGCAAGCCCTCTTACCGGAA---ACC 293465
Qy 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp 315
Db 293466 ACACAATCCACCAACCAATTTGGCGTACACGGGTAAAAACCTGGGCTTTGTGAAAAA 293525
Qy 316 AspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyr 335
Db 293526 GATGCCAACCGCTATGTGT---GGAAGAAAGACGCTATTCCGCGCATGACGCGGCAC 293581
Qy 336 GluLeuArgAlaGluTrpLys-----GlnProPhePro-----GlyPheGlu 349
Db 293582 CGGTTACCGACGCAATGTAAAGAGCGCCCAACCATACCCAAATCACCCTCGGGGTATGAA 293641
Qy 350 --AlaLeuArg-ValHisLeuAsnArgAsn-----AspTyrHis 361
Db 293642 CTTCAACTTCACACCGCCCTTCGCGAACAAACCTGCTGAATACGGTATCAACTACCG 293701
Qy 361 sHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAl 381
Db 293702 CCATCAGGAATCAAAACCGCAAGCG-----TTTTTGAATTCACAA-----TT 293743
Qy 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyr 401
Db 293744 TAAATTTGAATGAATAAGAAAGAACACTGATGAGAGAGAAATAAG----- 293789
Qy 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThr-SerGluAlaValLysGlnProMetLe 421
Db 293790 -----AACCGTGAANAATGAAAAAATGCAAAAGCTACCG 293824
Qy 421 uLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAs 440
Db 293825 TCTGACCAACCGCCAGCAAAACCGATACCGGCGGTATATCGAAGCATTACGAGATTGA 293884
Qy 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db 293885 CGGCTTTACCTGACCGCGGGCTGCGTTACGCGCTTCAAGGTGAAA---ACCCACGA 293941
Qy 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyValAla 480
Db 293942 CGGCATAAACCGTTTCAAGCAACAACCTTAAC-----CCGAGTTTCGGCGGTAT 293989
Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy 500

```

```

Db 293990 TTGGCAGCGGCACGAA-----CACTGGAGCTTTCAGCGCGAGCCACAA 294031
Qy 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 294032 CTAC-----GCCAGCGCGAGCCCGGCTG-----TATGACGGCTCTCAACCCCA 294076
Qy 520 sGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsnLysHisLe 536
Db 294077 CGCAAAACCGCGCATCATCTCGATTGCGGACGCGCAG----- 294113
Qy 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGl 556
Db 294114 -AAAGCCGAACCGCGCGCAATACCGAAATCGGCTTCACTACACGACGCGCACCTTTCG 294172
Qy 556 nTyrAsn-----LeuAlaLeuTyrArgAsnAr 565
Db 294173 CGCAAAACCGCAGCTACTTCTGGCAGACCATCAAGACCGGCTTGCCTATCCGCAAAACCG 294232
Qy 565 gPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 294233 CCACGACTCTGTCGCGCTCGGTGAAGCGCTCAATGCGGTTTACATCAAAAACCCAGGTTA 294292
Qy 579 -----ArgGl 580
Db 294293 CGAATGGCGCGCTCTTACCGCAGCGCGGCTGACTGCCAAAAGTCGCGTAAGCCACAG 294352
Qy 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db 294353 CAAACCGCGCTTTTACGATACGCACAAAGACAGCTGTGTAGCGCGAATCTCGAA----- 294407
Qy 600 aAspPheTyrGlyAlaGlu-----GlyGluIleTyrPheLysPr 613
Db 294408 -----TTTGGCGCACAAAGTCGCGCGCACTTGGACGCTTCCCTTGCCTACCGCTTCCA 294460
Qy 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLe 633
Db 294461 AAACCCGNAATCTGGAATTCGGCTGCGCGCGCGCTTATGTT----- 294500
Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 294501 -----CAAAAAGCGCTGGTTCGNATATTGGTGGCAGGTCAAAAAGA 294541
Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 294542 CCGCAACGCGCAAAATGGAAAAACGTTTACGCAAAAGGTTTCGGTGTGAAC----- 294590
Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 294591 -----GATGCTCTCCCAACTGGAAACCGCTGGG 294619
Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 294620 CAAAGACACGCTCAAT-----GTTAATCTTTCCGTTTAA----- 294653
Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerVa 733
Db 294654 -----AACGTGTTCAACACGTTCTA 294673
Qy 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
Db 294674 CTATCCGCAACGCCAACGATGGACCAATACCTCCGGGCGTGGACGCTGATGTACGCTT 294733
Qy 752 yGlyValAsnValLysPhe 758
Db 294734 GGGCGTGAACCTACAAGTTTC 294752

```

RESULT 29
 ID AAA81489
 XX AAA81489 standard; DNA; 837096 BP.
 AC AAA81489;
 XX
 DT 04-DEC-2000 (first entry)


```

QY 336 GluLeuArgAlaGluTrpLys-----GlnProPhePro-----GlyPheGlu 349
Db 658353 CGGTTACGAGCAATGTAAGAGCGCCCAACCATACCAATCACCACGCGGTATGAA 658412
QY 350 ---AlaLeuArg-ValHisLeuAsnArgAsn-----AspTyrHis 361
Db 658413 CTTCAACTTCGACGCGGCTTCCGGAACAAACCCCTGCTGAAATACGGTATCAACTACCG 658472
QY 361 sHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnGlnThrGlnAsnAl 381
Db 658473 CCATCAGGAATCAAAACCGCAAGCG-----TTTTTGAATTCAAA-----TT 658514
QY 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTy 401
Db 658515 TAAATTTGAGATAAAGAAAGAAACCACTGATGAGAGAAATAAG----- 658560
QY 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLe 421
Db 658561 -----AACCGTGAATGAAATAATGCCAAAGCCTACCG 658595
QY 421 uLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAs 440
Db 658596 TCTGACCAACCCGACCAACCAACCATACCGCGGGGTATATCGAAGCCATTCCAGAGATTGA 658655
QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db 658656 CGGCTTTACCTGACCGCGGGCTGCGTTACGACCGCTTCAAGGTGAAA---ACCACAGA 658712
QY 460 pLysAlaLeuLeuAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
Db 658713 CGCAAAACCGTTTCAAGCAACACCTTAAC-----CCGAGTTTTCGGGTGAT 658760
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLy 500
Db 658761 TTGCGCAGCCGACGAA-----CACTGGAGCTTCAGCGCGAGCCACAA 658802
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 658803 CTAC-----GCCAGCGCAGCCGCGCTG-----TAGCAGCGCTGCAACCCCA 658847
QY 520 sGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsnLysHisLe 536
Db 658848 CGGCAAAACGGCGCATCATCTCGATTGCGCAGCGCAGC----- 658884
QY 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGl 556
Db 658885 -AAAGCCGACGCGCGCAATACCGAATCGGCTTCAACTACAAACGCGCAGCTTGC 658943
QY 556 nTyrAsn-----LeuAlaLeuTyrArgAsnAr 565
Db 658944 CGCAACGGCAGCTACTTCTGGCAGACCATCAAGACGCGCTTGCCTCCAAATCCGCAAAACCG 659003
QY 565 gPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 659004 CCACGACTCTGTCGCGTCCGTCGTAAGCGGTCAATGCGGTTACATCAAAACCAACCGGTTA 659063
QY 579 -----ArgGl 580
Db 659064 CGAATTGGGGCGCTCTACCGCACCGGCGGCGCTGACTGCGCAAGTCGGCGTAAGCCACAG 659123
QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db 659124 CAACCGCGCTTTTACGATACGACAAAGACAGCTGTTGAGCGCGGATCTCTGAA----- 659178
QY 600 aAspPheTyrGlyAlaGlu-----GlyGluIleTyrPheLysPr 613
Db 659179 -----TTTGGCGCACACAGTCGCGCGCACTTGACGCGCTTCCCTTACCGCTTCCA 659231
QY 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyLeuLysAsnLe 633
Db 659232 AAACCGGAATCTGGAATCGCTGCGCGCGCGCTTATGTT----- 659271
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653

```

```

Db 659272 -----CAAAGCCGCGGTTCGATATTGGTGGCAGGTCAAAAGA 659312
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 659313 CGCAACGCAATTCGAAACGTTGTACGCAAGGTTTCGGTGTGAAC----- 659361
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 659362 -----GATGTCTTCGCCCAACTGGAAACCGCTGGG 659390
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 659391 CAAAGCACGCTCAAT-----GTTAATCTTTCGGTTAAC----- 659424
QY 713 nThrArgTyrGlyGluTyrAsnTyrValLysAlaAspAsnLeuLeuAsnGlnSerVa 733
Db 659425 -----AACGTGTTCAACAGTTCTA 659444
QY 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
Db 659445 CTATCGCACAGCAACGATGGACCAATACCTCGCGGCGTGGACGTGATGTACGCTT 659504
QY 752 yGlyValAsnValLysPhe 758
Db 659505 GCGCGTGAACATACAGTTC 659523

```

RESULT 30
ABQ78300
ID ABQ78300 standard; DNA; 2307 BP.
XX ABQ78300;
XX 05-NOV-2002 (first entry)
XX Nucleotide sequence of p64 polypeptide.
XX p177; p88; p64; p55; p46; vaccine; gonorrhea; gene; ss.
XX Neisseria gonorrhea.
XX Key Location/Qualifiers
XX CDS 99..2261
XX FT /*tag= a
XX FT /product= "p64"
XX PN WC200260936-A2.
XX PD 08-AUG-2002.
XX PF 31-JAN-2002; 2002WO-US02881.
XX PR 31-JAN-2001; 2001US-266070P.
XX PR 06-AUG-2001; 2001US-310356P.
XX PR 23-OCT-2001; 2001US-344432P.
XX XX (IOWA) UNIV IOWA RES FOUND.
XX PA (REGC) UNIV CALIFORNIA.
XX PA (APIC/) APICELLA M A.
XX PA (EDWA/) EDWARDS J L.
XX PA (GIBS/) GIBSON B W.
XX PA (SCHE/) SCHEFFLER K.
XX PA (BROW/) BROWN E.
XX XX Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
XX WPI; 2002-619227/66.
XX DR P-PSDB; ABB78069.
XX XX New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
XX PT gonorrhea, useful for preventing, or protecting a female patient
XX PT against, N. gonorrhea colonization or infection -


```
Db 1620 -----AAAGCCGAACGGCGCGCAATACCGAATAATCGGCTTCAA 1657
Qy 549 yTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
Db 1658 CTACACAGCAGCGCAGCTTTGCGCAAC-----GGCAGCTA 1693
Qy 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerG 589
Db 1694 CTTCTGG---CAGACCATCAAGACGCGCTTGCCTACCGCAAAACCGCAGCACTCTGT 1750
Qy 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyG 608
Db 1751 CCGCGTCCGTGAAGCGCTCAATGCCCGGTACATCAAAACCCAGGTACGATTATGGGCGC 1810
Qy 608 uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG 628
Db 1811 GTCCTACCGCAGCGCGGCGCTGACTGCCAAAGTCGCGGTGAGC-----1853
Qy 628 yArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
Db 1854 -----CACAGCAAAACCGCGCTTTTACGATACGCACAAAGACAAGCTGTT 1897
Qy 648 eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db 1898 GAGCGCGAATCTGAATTT-----GGCGCACAAAGTCGCGCACTTG 1939
Qy 668 uLysAlaSerLeuThrAspArgIle---AspAlaAsnLeuAspTyr---TyrArgValPh 686
Db 1940 GACGGCTCCCTTGCCTACCGCTTCCAAATCGGAATCTGGAATCGGCTGGCGC-----1994
Qy 686 eAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLe 706
Db 1995 -----GGCGGTATTGTTCAAAAAGCTACGGGT---TCGATATTGGCGGC 2035
Qy 706 uGlyAlaAsnTyrArgArg-----As 713
Db 2036 AGTCAAAAAGACCGCAAGCAACTTGGAAACGTTGTACGCAAAAGTTTCGGTGTGAA 2095
Qy 713 nThrArgTyrGlyGluTrp-----AsnTyrTyrValLysAl 725
Db 2096 CGATGCTTTCGCCCAACTGGAAACCGCTGGGCAAAAGACACGCTCAATGTCAATCTTTCGGT 2155
Qy 725 aAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---Pr 744
Db 2156 TACACAGCTGTTCAACAGTTCTACTATCGCACAGCCCAACGCTGGACCAATACCTGCC 2215
Qy 744 oGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2216 GGGCGTGGGACGTGATGATACGCTTGGCGTGAACCTACAAGTTC 2258
RESULT 31
AAF26411
ID AAF26411 standard; DNA; 2586 BP.
XX AC
XX AAF26411,
XX
XX
XX 02-MAY-2001 (first entry)
DE
DE Pseudomonas sp heavy metal transporter encoding DNA ORF09606.
XX
XX Heavy metal transporter; iron transporter; transgenic plant;
KW homeostasis regulator; heavy metal ion; trace element; soil pollution;
KW plant growth promoter; plant development; ds.
XX
XX Pseudomonas sp.
OS
XX
XX DE19934720-A1.
XX
XX 25-JAN-2001.
XX
XX 23-JUL-1999; 99DE-1034720.
XX
XX 23-JUL-1999; 99DE-1034720.
```

```
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
DR WPI; 2001-160596/17.
PT
PT New DNA encoding iron or heavy metal transporters, useful for promoting
PT growth of plants on polluted soil -
XX
PS Claim 3a; Page 7-8; 54pp; German.
XX
CC This invention describes novel DNA sequences (I) that encode iron or
CC heavy metal transporters (II). The invention also describes (1) a
CC recombinant expression vector (III) containing (I); (2) prokaryotic or
CC eukaryotic cells (III) transformed with (I) or (III); (3) production of
CC (II) by culturing (III); (4) (partial) expression products (IV) of (I)
CC and synthetic proteins or peptides with the same sequences; (5)
CC antibodies (Ab) specific for (IV); (6) hybridoma cells that produce
CC monoclonal Ab; and (7) transgenic plants that contain (III). The iron or
CC heavy metal transporters encoded by (I) transport heavy metal ions across
CC cell walls and regulate homeostasis of trace elements. (I), and their
CC fragments are useful for: (1) expression of (II); (2) as probes and
CC primers for detection, isolation and amplification of full length cDNA
CC sequences; and (3) producing transgenic plants. (III) are used to promote
CC growth, development and yield of plants, particularly leguminosae,
CC especially when growing in soil polluted by heavy metals injurious to
CC plants. They also improve homeostasis of iron and trace elements. Host
CC cells that express the iron or heavy metal transporters take up heavy
CC metals, so reduce pollution of soil and release iron and other trace
CC metals which improves soil quality and protect plants against pollutants
CC from the soil.
XX
SQ Sequence 2586 BP; 563 A; 857 C; 742 G; 421 T; 3 other;
```

Alignment Scores:

```
Pred. No.: 3,31e-10 Length: 2586
Score: 226.50 Matches: 174
Percent Similarity: 36.34% Conservative: 132
Best Local Similarity: 20.67% Mismatches: 336
Query Match: 5.61% Indels: 201
DB: 22 Gaps: 39
```

US-09-936-377-2 (1-758) x AAF26411 (1-2586)

```
Qy 17 AsnThrProLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrVal 36
Db 346 AACGGCAATGTGTGTACTGGAGCGCCAAACCGCAAGCAACGTGATCCCTCGACGAGGTG 405
Qy 37 ThrValValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAla 56
Db 406 ACCTGACGCGCCACCGCGACGACAGGACGTCGACGAGGTG-----CCGACACTGTC 459
Qy 57 SerAspLysIleLeuSerGlyAspThrLeu---ArgGlnLysAlaValAsnLeuGlyAsp 75
Db 460 AGC-----GTGCAGACCCGCTGACGCAACTGGACCGCAGAACCGTGAACATATCCAGGAT 513
Qy 76 AlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGly-----GlyGlyAlaSer 93
Db 514 CTGGTCCGTACAGCCGGGGTTCGGTATCCGGAACCGCGCAACCGCGCGCTGTAAC 573
Qy 94 AlaProValIleArgGlyGlnThrGlyArgAlaIleLysValLeuAsnHisHisGlyGlu 113
Db 574 GGCTACACATCCGTGTATCGACGCGGAGCGGATCCTCACCACAGTCGCGCGGTGTCG 633
Qy 114 ThrGlyAspMetAlaAspPheSerPro-----AspHisAlaIleMetValAspThr 130
Db 634 ATCCCGCAGAGCTTCTTCTACGCGCCCTTACGCGCCCTACGCGCAGCGCACTACGCGCGC 693
Qy 131 AlaLeuSerGlnGlnValGluLeuArgGlyProValThrLeuLeuTyrSerSerGly 150
```


QY	407	rAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGI	427	QY	708	lAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnL	728
Db	1143	TAAAAAAGATGCAACTGAGGAAGATAAAAGAGAACCGTGAAATGAAAAATGCCAA	1202	Db	2036	TTAAC	2044
QY	427	nHisTyrSerPhePheGlyValGluGlnAlaAsnTrp	439	QY	728	eULeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr	747
Db	1203	AGCCTACCGTCTGACCAACCGACCAACACCATACCGCGGTATATCGAAGCCATTCA	1262	Db	2045	TGTTTCGACAGTCTACTATCCGACGACCAAGCTGGACCATCCCTCGCGGGCTGG	2104
QY	440	-----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerI1	457	QY	747	lyArgSerPheThrGlyGlyValAsnValLysPhe	758
Db	1263	CGAGATTACGGCTTTACCTGACCGCGGCTGCTTACGACCGCTTCAAGTGAACAC	1322	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	457	eArgTyrAspLysAlaLeuLeuAspArgGluAsnTyrTyrLys	476	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1323	COAC	1363	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	476	pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT	496	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1364	GTTC	1405	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	496	hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG	516	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1406	GCAGGAGCCCAACTAC	1450	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	516	luLeuTyrAlaHisGlyLysHis	532	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1451	CGCTGCAACACCGGCAACCGCGGATCATCTCGATTCCGACGGCAC	1500	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	532	lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG	552	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1501	-----AAAGCCGAGCGCGCGCAATACCAATCGCTTCACTACACG	1546	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	552	lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg	563	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1547	ACGCGACGCTTCCGCAACCGCGGATCATCTCGGACGACCATCAAGACGGCTTGCA	1606	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	564	-----AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly	578	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1607	ATCCGCAAAACCGCAGCACTGTGCGCGTCCGGAAGCCGTCAACGCGGCTACATCA	1666	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	578	-----	578	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1667	AAACACCGTTCAGATTTGGCGGTCTCTCCGACCGCGCGCTGACCGCCAAAGTCG	1726	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	579	-----ArgGlyProLysSerIleGluAspAspSerGluMetLysLeuVala-gTyrA	596	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1727	CGGTAAGCGCGCAACACCGCGCTTTACGATACCCATCCTTAAACAACTGTTGAGCGCA	1786	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	596	snGlnSerGlyAlaAspPheTyrGlyAlaGlu--GlyGluIleTyr	610	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1787	ACCCGAG	1834	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	611	-----PheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG	628	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1835	CCTACCGCTTCAAA	1887	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	628	lyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProp	648	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1898	-----CNAACCTACGGGTTCGATATTGG	1912	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	648	heIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL	668	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1913	CGGAGGGCAAAAGACCGCGCAATTTGGAACACCTGTAGCCAGGTTTCGGTG	1972	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	668	eUlyAlaSerLeuThrAspArgIleAlaAlaAsnLeuAspTyrTyrArgValPheAlaG	688	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1973	TGAAC	1990	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
QY	688	InAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyA	708	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139
Db	1991	ACTGGAACCGCTGGGCAAAAGACACGCTCAAT	2035	Db	2105	GACGTGATGTCACGCTCGCGCGTGAACACTACAAGTTC	2139

Alignment Scores:

Pred. No.:	6,05e-10	Length:	2368
Score:	223.00	Matches:	193
Percent Similarity:	33.45%	Conservative:	99

RESULT 33

AAAT42128	standard; DNA; 2368 BP.
XX	AAAT42128;
XX	02-MAR-1998 (first entry)
DE	Neisseria gonorrhoeae strain FA1090 frpB gene.
KW	FrpB; outer-membrane protein; iron regulating; immunogen;
KW	vaccine; diagnosis; infection; Neisseria gonorrhoeae;
KW	Neisseria meningitidis; ss.
OS	Neisseria gonorrhoeae.
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	160..2302
FT	/tag= a
FT	/product= frpB
FT	/note= "70 kD major iron-regulated outer-membrane protein"
PN	WO9631618-Al.
XX	10-OCT-1996.
XX	08-APR-1996; 96WO-US04774.
XX	07-APR-1995; 95US-0418964.
XX	(UYNC-) UNIV NORTH CAROLINA.
XX	Beucher M, Sparling PF;
XX	WPI; 1996-465038/46.
XX	P-PSDB; AA05181.
XX	DNA encoding Neisseria iron-regulated outer membrane protein FrpB -
XX	useful for producing vaccine to protect mammals against N.
XX	gonorrhoeae and N. meningitidis infection
XX	Claim 3; Fig 10; 68pp; English.
XX	This sequence encodes a novel iron-regulated outer-membrane protein
XX	FrpB which was isolated from the gonococcal microorganism Neisseria
XX	gonorrhoeae strain FA1090. This gene is also homologous to one found in
XX	Neisseria meningitidis. The FrpB protein is surface exposed and
XX	immunogenic and is predicted to be useful as a vaccine because of its
XX	surface exposure, partial antigenic conservation and susceptibility to
XX	attack by bacterial antibodies. Such a vaccine will be of use against
XX	infection in mammals by N. gonorrhoeae and N. meningitidis. The protein
XX	and its fragments will also be useful in the diagnosis of Neisseria
XX	type infections.
SQ	Sequence 2368 BP; 685 A; 709 C; 542 G; 432 T; 0 other;

Best Local Similarity:	22.11%	Mismatches:	291
Query Match:	5.53%	Indels:	292
DB:	17	Gaps:	45
 US-09-936-377-2 (1-758) x AAT42128 (1-2368)			
Qy	11	LeuSerIleLeuLeuIleAsnThrProLeu--LeuAlaGlnAlaHisGluThrGluGln	29
Dd	180	CTCAGCCCTGCCTCGCTCACACTTCGCCGGCTTTGCCACCGCGCAAAATAATGCC	239
Qy	30	SerValGlyLeuGluThrValThrValValGlyLyysSerArgProArgAlaThrSerGly	49
Dd	240	AATGTCCGATTGGATACCGTACCCTTAAGAGCGACCGCAA-----	281
Qy	50	LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrIleuArgGlnLys	69
Dd	282	-----GGCAGCAAAATCCGTACCAACATCGTT-----ACGCTTCAACAAAAA	323
Qy	70	-----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla	85
Dd	324	GACGAAGACCGCAACCGATATGCGCGAACCTCTTAAAAGAAGACCCCTCCATC-	377
Qy	86	SerGlnTyrrGlyGlyAlaSerAlaProValIle-----ArgGlyGln	100
Dd	378	---GATTTCGCGCGCGCAACCGCACGCTCCCAATTCCTGACGCTCGCGGCGATGGTGCAG	434
Qy	101	ThrGlyArgArgIleLysVal-----LeuAsnHisHis	111
Dd	435	AACCTCTGTCGACATCAAGSTGGACACCGCTATTCGACAGCCAAATCTTTACCACCAA	494
Qy	112	GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValaspThrAla	131
Dd	495	GGCAGA-----TTTATTGTGATCCCGCT	518
Qy	132	LeuSerGlnGlnValGluIleuArgGlyProValThrLeuLeuTyrrSerSerGlyAsn	151
Dd	519	TTGTTAAAGTCGTTTCGTACAAAAGGC-----CGGGGTGCC	557
Qy	152	ValAlaGlyLeuValAspValalaAspGlyLysIleProGluLysMetProGlu-----	169
Dd	558	GCCTCTGCGGTATCGCGCGCACCAACGCGCGGATTATCGCCAAACCGTCGATGCCCAA	617
Qy	170	-----AnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu	186
Dd	618	GACCTGTCAAAGGCTTCGATAAAACTGGGCGGTGCGCTCAACAGCGGC-----	668
Qy	187	LysLeuThrSerGlyGlyIleAsnIledgyl-----LeuGlyLys-----AsnPhe	201
Dd	669	TTTGCCGCAACAACGCGGTAAAGCTACGCGCAACGCTATTCGMAAAGAGGCACTTC	728
Qy	202	ValLeuHisThrGluGlyLeuTyrr-----ArgLysSerGlyAspTyrrAlaVal	217
Dd	729	-----GACGGFTTGTCTCTTACACCGCAACGATGAAAAAGATTACGAAGCC	776
Qy	218	ProArg-----TyrArgAsnLeu-----LysArgLeuProAspSerPro-----	230
Dd	777	GGCAAGGCTTCGCAATGTCAACGCGCGCAAAACCGTACGTCACGCGCGCTGCACAAA	836
Qy	231	-----ArgArgPheAlaAsnGlyGlnHisArgAlaValLeu	242
Dd	837	CGCAGCTACCTCGCCAAATCGGAACAACCTTCGGCGACGCGCACCGCATCGTATTG	896
Qy	243	GlyTrpArgLysArgPheTyrrArg-----ArgThrTyrrSerAspArgAspGlnTyrr	260
Dd	897	AGCATATGAAGACCACACCGGGGCGATCGCACTGTG-----CGTGAAGAGTTT	947
Qy	261	GlyLeuProAlaHisSerHisGluTyrrAspAspCysHisAlaAspIleIleTyrrGlnLys	280
Dd	948	GCCGTCGCGCGCAAAAT-----TCA	968
Qy	281	SerLeuIleAsnLysArgTyrrLeuGlnLeuTyrrProHisLeuLeuThrGluGluAspVal	300
Dd	969	CGGATAACTATAACAGCCAAAGCCCTCCCTACCGCGAA---ACCACACAATCCAACACC	1025

QY 596 snGlnSerGlyAlaaspPheTyrGlyAlaGlu---GlyGluIleTyr----- 610
 Db 1945 ACCCGAG-----TTGGCGCACAAACCGCGCGCACTTGGAGCGGCTCCCTTG 1992
 QY 611 -----PhelysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG 628
 Db 1993 CCTACCGCTTCAAA---AACCCGAATCTGMAATCGCTGGCGGACGCTATGTT--- 2045
 QY 628 lyArgLeuLysAsnLeuProSerLeuProGlyArgGluaspAlaTyrGlyAsnArgProp 648
 Db 2046 -----CAAAAAGCTACGGGTTCGATATTGG 2070
 QY 648 heileAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL 668
 Db 2071 CGCAGGGCAAAAGACCGCGCAANTTGGAAAACGTTGTACGCCAAGGTTTCGGTG 2130
 QY 668 euLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG 688
 Db 2131 TGAAC-----GATGCTTCGCCA 2148
 QY 688 lnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyA 708
 Db 2149 ACTGGAACCGCTGGCAAGACACGCTCAAT-----GTTAATCTTTCGG 2193
 QY 708 laAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnL 728
 Db 2194 TTAAC-----AACG 2202
 QY 728 euLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetG 747
 Db 2203 TGTTCACAAAGTTCTACTATCCGCACAGCAACGCTGGACCAATACCTCGCGGGGTG 2262
 QY 747 lyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2263 GACGTGATGTCGCTGGCGGTGAACACTACAAGTTC 2297

RESULT 34
 AAD10202
 ID AAD10202 standard; DNA; 2381 BP.
 XX AAD10202;
 XX 24-SEP-2001 (first entry)
 DE N. gonorrhoeae or N. meningitidis strain FA1090 FrpB protein DNA.
 XX Fe-regulated protein B; FrpB; outer-membrane protein; vaccine; infection;
 KW antibacterial; ds.
 XX Neisseria gonorrhoeae.
 OS Neisseria meningitidis.
 XX
 FH Key Location/Qualifiers
 CDS 172..2313
 FT /*tag= a
 FT /product= "N. gonorrhoeae or N. meningitidis
 FT strain FA1090 FrpB protein"
 XX
 PN US6265567-B1.
 XX
 PD 24-JUL-2001.
 XX
 PF 05-APR-1996; 96US-0628434.
 XX
 PR 07-APR-1995; 95US-0418964.
 XX
 PA (UYNC-) UNIV NORTH CAROLINA.
 XX
 PI Sparling PF, Beucher M;
 XX
 DR WPI; 2001-450739/48.
 DR P-PSDB; AAE05408.
 XX

PT An isolated nucleic acid encoding a Fe-regulated protein B (FrpB),
 PT useful for production of a vaccine against infection by Neisseria
 PT gonorrhoeae or N. meningitidis -
 XX
 XX Claim 2; Column 29-34; 20pp; English.
 XX
 CC The present DNA sequence encodes Fe-regulated protein B (FrpB) isolated
 CC from Neisseria gonorrhoeae and Neisseria meningitidis (strain FA1090). The
 CC present DNA sequence also represents the combined nucleotide sequence
 CC from PUNCH319 and PUNCH325. The FrpB is a 70 kD major iron-regulated,
 CC outer-membrane protein common to N. gonorrhoeae or N. meningitidis.
 CC The FrpB has particular use as a vaccine that protects a mammal from
 CC N. gonorrhoeae or N. meningitidis infection. The FrpB may also be used
 CC to detect the presence of antibodies specific for N. gonorrhoeae or
 CC N. meningitidis in a sample.
 XX
 SQ Sequence 2381 BP; 690 A; 712 C; 543 G; 436 T; 0 other;
 Alignment Scores:
 Pred. No.: 6.09e-10 Length: 2381
 Score: 223.00 Matches: 193
 Percent Similarity: 33.45% Conservative: 99
 Best Local Similarity: 22.11% Mismatches: 291
 Query Match: 5.53% Indels: 292
 DB: 22 Gaps: 45

US-09-936-377-2 (1-758) x AAD10202 (1-2381)

QY 11 LeuSerIleLeuLeuIleAsnThrProLeu---LeuAlaGlnAlaHisGluThrGluGln 29
 Db 193 CTCAGCTCTCTCGCTCACACTTGCCTGGCGGCTTTCACCGCGGCAAAATAATGCC 252
 QY 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49
 Db 253 AATGTCGATTGATACCTTACCGTAAAGGCGCGCCAA----- 294
 QY 50 LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLys 69
 Db 295 -----GGCAGCAAAATCCGTACCAACATCGTT-----ACGCTTCAACAAAA 336
 QY 70 -----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
 Db 337 GACGAAGACCGCAACCGATATGCGGAATCTTTAAAGAGAGAGCCCTCCATC----- 390
 QY 86 SerGlnTyrGlyGlyGlyAlaSerAlaProValIle-----ArgGlyGln 100
 Db 391 ---GATTTCGGCGGCGCAACGCGACGCTCCCAATTCCTGACGCTGCGCGCATGGGTCAG 447
 QY 101 ThrGlyArgArgIleLysVal-----LeuAsnHisHis 111
 Db 448 AACTCTGTCGACATCAAGGTGGACACACGCTATTCCGACAGCCAAATCCTTTACCACCAA 507
 QY 112 GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla 131
 Db 508 GGCAGA-----TTTATTGTCGATCCCGCT 531
 QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
 Db 532 TTGGTTAAAGTCGTTTCGCTACAAAAGGC-----GCGGTTC 570
 QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu----- 169
 Db 571 GCCTCTGCGGTATCGCGCGCAACACGCGCGATTTATCGCCAAAACCGTCGATGCCCAA 630
 QY 170 -----AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186
 Db 631 GACCTGCTCAAGGCTTGGATPAAACTGGGCGGTGGCCTCAACAGCGC----- 681
 QY 187 LysLeuThrSerGlyGlyIleAsnIleGly-----LeuGlyLys-----AsnPhe 201
 Db 682 TTTGCGGCAACACACGCGGTAAAGTACGCGCAAGGATTTCGGAAGAGGGCACTTC 741
 QY 202 ValLeuHisThrGluGlyLeuTyr-----ArgLysSerGlyAspTyrAlaVal 217

XX WO200259320-A2.
 XX 01-AUG-2002.
 XX 19-OCT-2001; 2001WO-US46833.
 XX 19-OCT-2000; 2000US-242412P.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blattner FR, Welch RA, Burland VD;
 XX WPI; 2002-691532/74.
 XX New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
 XX useful for preventing or treating E. coli CFT073 infection in humans or
 XX livestock -
 XX Claim 1; Page 166-184; 765pp; English.
 XX The present invention relates to polynucleotide sequences from the
 XX genome of the pathogenic Escherichia coli strain CFT073. Almost all
 XX the sequences present in E. coli CFT073 are absent in the previously
 XX sequenced laboratory strain K-12. The polynucleotide sequences of
 XX the invention are useful for preventing, diagnosing or treating
 XX E. coli CFT073 infection in humans or livestock. The polynucleotide
 XX sequences are useful for preventing urinary tract infections and
 XX pyelonephritis. Likewise, the polypeptides encoded by the different
 XX open reading frames (ORF1-5) are useful for generating a vaccine
 XX against uropathogenic E. coli strains. ABS78834-ABS79085 represent
 XX genomic sequences from E. coli strain CFT073.
 XX SQ Sequence 32160 BP; 7001 A; 9686 C; 8766 G; 6688 T; 19 other;
 Alignment Scores:
 Pred. No.: 3,71e-08 Length: 32160
 Score: 220.50 Matches: 187
 Percent Similarity: 34.05% Conservative: 97
 Best Local Similarity: 22.42% Mismatches: 309
 Query Match: 5.46% Indels: 242
 DB: 24 Gaps: 45
 US-09-936-377-2 (1-758) x ABS78887 (1-32160)
 QY 1 MetAlaGlnThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuLeuLeuLeu 20
 DB 29138 ATGAAATACACACGGCTTTATCTCTGGCTTGGGGGATATTGCTC-----CCCGCC 29191
 QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
 DB 29192 ATTGCTAATGCCAGACTTCACAGCAGAC-----GAAGCAGCGCTGGGTATCC 29242
 QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 DB 29243 GCCAGTAACAATCTTCGGCTCGGCA-----TCAGCCACAACAGCTCTCATCTCTGT 29296
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 DB 29297 GTCAGCGCCCGGAATTAACGACGACCGCGCTCACCGCAGGACAAACTCCCGAGATC 29356
 QY 81 ---ProGlyIleHisAlaSerGlnTyGlyGlyGlyAlaSerAlaProVal---IleArg 98
 DB 29357 TTGCCCGGGCTCAATATTGAAATACGCGCAACATGCTTTTTCGAGATCTCGCTACGC 29416
 QY 99 GlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAla 118
 DB 29417 GCGGTCTCTTCAGCGCAG----- 29434
 QY 119 AspPhe---SerProAspHisAlaIleMetValAsp----- 129
 DB 29435 GACTTCTATACCCCGCGCTCACCCCTGTATGTCGATGGCGTCCCTCAGCTTCCACCAAC 29494

QY 130 -----ThrAlaLeuSer-----GlnGlnValGluIleLeuArgGlyProValThrIleu 145
 DB 29495 ACCATCCAGCGCGCTTACCGATGTGCAAGCGTGTGAGTTGCTGCGAGGCCACACGGAACG 29554
 QY 146 LeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGlu 165
 DB 29555 TTATATGGCAAAAGCGCTCAGCGCGGATCATCAATCGTCACCCAGCAG---CCGGAC 29611
 QY 166 LysMetPro-----GluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSer 181
 DB 29612 AGCACGCGCGCGCTATATTGAAGCGCGCTCAGTAGCCGACAGTTATCGA---AGT 29668
 QY 182 SerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPhe 201
 DB 29669 AAGTTCAACCTG-----AGCGGCCCATTCAGGATGCCTG----- 29704
 QY 202 ValLeuHisThrGluGlyLeuTyrArgLys-----SerGlyAspTyrAlaValProArg 219
 DB 29705 CTGTACGCGACGCTCACCTGTTACGCCAGGTTGATCAGCGGCACATGATTAA-CCCGCG 29763
 QY 220 -----TyrArgAsnLeuLysArgLeuProAspSerProArg-----ArgPheAlaAsn 235
 DB 29764 GACGGGAGCGATGACTTAGCGGCGCACCGCGCAGCATAGGAATGAACTGCGTCT 29823
 QY 236 GlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAsp 255
 DB 29824 GCGCGCGACGATCAGCCCTGGCA----- 29847
 QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAsp 275
 DB 29848 -----AATGGGCTT-----TGC----- 29859
 QY 276 IleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
 DB 29860 -----CGCTCAGCGAATGATACCCGCG----- 29883
 QY 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys-GlyPheHisAspAspAs 315
 DB 29884 -----CACCCAGGATGCTTATGTGGATGGAATGATATTAA 29919
 QY 315 pAspAlaHisAlaHisAlaHisAsnGlyLysPro----- 326
 DB 29920 GGGCCGTAAGCTGTGATCAGCGATGTTTACCACAGACCCGTATACGCGGCTGCACTGA 29979
 QY 327 -----TyrIleAspLeuArgAsnLysAr 334
 DB 29980 CAGCCAGACCTCAGTGGGAAATACACACCGATGACTGGGT----- 30022
 QY 334 gTyrGluLeuArgAlaGluTyrLysGln-----ProPheProGlyPheG1 349
 DB 30023 -TTCAACCTGTATCAGCGCTGGCAGCAGCAGCATATTTCGCGCACCTTCCCTTCC---GG 30078
 QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLys----- 365
 DB 30079 TTCGTTAATCGTCATATCCCTCAGCGC---TGGAAATCAGGATGTGCAGAGCTGCGCGC 30135
 QY 366 -----AlaGlyAspAla-----ValGluAsnPhePheAsnGlnThrGlnAsnAl 381
 DB 30136 CGCAACCTCGGCGATGCGCTACCGCTTATATGTTTGGCTGTACCGCAGAACAC 30195
 QY 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTy 401
 DB 30196 CCGC-----GAGAAGTAAATTCAAGCTACGACATGCGGAC 30231
 QY 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetIle 421
 DB 30232 AATGCTTATTAAACAGTACCGGCTATACCCAGCTGAACGCTGGCGCA----- 30283
 QY 421 uLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp----- 439
 DB 30284 -----TACAGT-----GACCTGACCTGGCATTT 30306
 QY 440 ----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleAr 458

Db 2113 AGTGTACCTACCAAGTTATATACTGACCGAATCTTGCTGCTACTGATGCG-----CCA 2166

Qy 715 ArgTyrGlyGluTrp 719
 |||||
 Db 2167 TTTAAAGCGCAATGG 2181

RESULT 37
 AAZ54329
 ID AAZ54329 standard; DNA; 2112 BP.
 XX AC
 XX AAZ54329;
 XX
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 761 partial DNA sequence SEQ ID NO:2607.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy; ds.
 XX
 XX Neisseria meningitidis.
 XX OS
 XX Neisseria meningitidis.
 XX
 XX
 EN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 XX
 XX 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 XX (GENO-) INST GENOMIC RES.
 XX
 XX
 PI Paterson C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 XX WPI; 2000-062150/05.
 DR P-PSDB; AAY75567.
 XX
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 XX
 PS Claim 7; Page 1236-1237; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAY75941
 CC represent novel *Neisseria* meningitis and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisserial* bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 2112 BP; 555 A; 650 C; 484 G; 423 T; 0 other;

Alignment Scores:
 Pred. No.: 4 53e-09 Length: 2112
 Score: 218.50 Matches: 182
 Percent Similarity: 33.98% Conservative: 117
 Best Local Similarity: 20.6% Mismatches: 260

```

QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAla 319
Db : : : : :
880 CTTGAATAC----- 888
QY 320 HisAlaHisAsnGlyLysProTrrPilleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db : : : : :
889 ---GCCTTCAACGACAAA---TGG-----CGTGCC 912
QY 340 GluTrrPlysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db : : : : :
913 CAATGGCAG-----CTGCCCCACCGCAGCGCGGCAGGAT 948
QY 360 TyrHisHis-----AspGluLysAlaGlyAspAlaValGluAsnPhenPhe----- 374
Db : : : : :
949 TTTGATCATTTCTATGTCAGGACGAGCAAAATGGCAACTTATCAAAAGTAACTACGCGCTGG 1008
QY 375 -----AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg---HisGlnProIle 390
Db : : : : :
1009 CAGCAGACCGACACAAACCGCTGCTGCCAAGCTTAACGCTCAACGGCGACTACACCAATC 1068
QY 391 GlyArgLeuLysGly-----SerTrrPglyValGlnTyrLeuGlyGlnLys----- 405
Db : : : : :
1069 GCGCGTTTGAACACCAACCTGACCGTAGGCATGGATTACAGCGCGCAACACCGCAACCCCG 1128
QY 406 -----SerSerAlaLeuSerAlaThr----- 412
Db : : : : :
1129 ACATGGGTTTCACGACGCGCTTTTCGCGCTTCATCAACCCCTACGACCGCGCAAGCTGG 1188
QY 413 ---SergluAlaValLysGlnProMetLeuLeuAspAsn-----LysValGlnHisTyr 429
Db : : : : :
1189 CCGGCTTCGGCGCAGATTGCGACCTTATCTGACCCCAAAACCGCACAAAGCGCACTCCTAC 1248
QY 430 SerPhePhe-----GlyValGluGluAlaAsnTrpAsp---AsnPherThrLeuGluGly 446
Db : : : : :
1249 GGCATCTTTGTGCAAAACATCTTCTCCGCGACGCGCGGATTGAAATTCGCTCCGCGCGC 1308
QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLysAspArg 466
Db : : : : :
1309 -----CGTTACGACAAATACACCTTTAATTC 1335
QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---ThrAlaArg 485
Db : : : : :
1336 GAAACAAACATC-----ACCGCGCAGCAGCGCCCAATACAGCGGACAC 1377
QY 486 SerPheAla-----LeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db : : : : :
1378 TCGTTACCCCAACATCGCGCAGTGTGGACATCAATCCCGTCCAC-----ACACTT 1431
QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db : : : : :
1432 TACGCTCGTATAACAAA-----GGCTTCGCGCTTATGCGGACGCGCGGCTAT 1482
QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db : : : : :
1493 TTGACATCATATAGTTGCTCTCCGCGGTGTTCAACCGCCGACCCCGAGTACACCGCCAA 1542
QY 544 IleGluLeuAlaLeu-----GlyTyrGluGlyAspArgTrrPgnTyrAsnLeuAlaLeu 561
Db : : : : :
1543 TACGAACCGCGGTGAAAGCAGTGTGGTGGACGACGCGCTCAGCACTACGTTGCTGCC 1602
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db : : : : :
1603 TAC----- 1605
QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
Db : : : : :
1606 -----CAAATCGACGCTTCAT----- 1623
QY 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArg-----TyrArg 618
Db : : : : :
1624 -----ATCGGCTACCGCGCGGATCCAAAAACCAACCCCTATATT 1662
QY 619 IleGlyValSerGlyAsp----- 624

```

```

Db 1663 TATGCGGTTAGCGGCAAAACACCGTTCCGCGCGGTGGAATTTGTCGCCCATCGGCAATC 1722
QY : : : : :
625 -----TyrValArgGlyValArgLeuLysAsnLeuProSerLeuProGlyArg 639
Db : : : : :
1723 ATCCCAAAAAAATCTATCTGCGCGGTTCGTTGGCGGTGATGACGCGGAAA----- 1773
QY 640 GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
Db : : : : :
1774 -----GTCGTTGAAGACAAGAAAAATCCCGACCGAGTG 1806
QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db : : : : :
1807 -----GGCATCCATTG-----ATAATACCAACGTTACCGGCAAC 1845
QY 680 LeuAspTyrTyrArgVal----- 685
Db : : : : :
1846 CTG---TTTTCGTTATATACCCCGACGAAACCTCTACGCGAAATCGCGTAACCGGT 1902
QY 686 -----PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
Db : : : : :
1903 ACAGCAAAACGCTACCGTTACAACTCAAGAAATAAAGAGTGACTACGCTTCCAGGCTTT 1962
QY 702 HisMetLeuAsn-----LeuGlyAlaAsnTyrArgArg---AsnThrArgTyrGlyGlu 718
Db : : : : :
1963 GCGCGAGTTGATGCCATCTGCTGGCTGGAACCAATAAAATGTTACGTTACCTTGCCTCC 2019
QY 719 TrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSer 738
Db : : : : :
2020 -----CGACCAATCTGCTCAATCAAAATAATTTGGCGTTTCGAGACTCT 2061
QY 739 PheLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db : : : : :
2062 ATG-----CCGGTAAATCCGCGCTATATCTGCCCGGTAAATTACCGTTTC 2109

RESULT 38
AAF56467
ID AAF56467 standard; DNA; 2112 BP.
XX
AC AAF56467;
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis coding sequence #26.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
XX rni5; rth; tolC; ds.
XX
OS Neisseria meningitidis.
XX
PN EP1069133-A1.
XX
PD 17-JAN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nassif X, Tinsley C;
XX
DR WPI; 2001-082916/10.
XX P-PSDB; AAB68927.
XX
PT Immunogenic polypeptides derived from Neisseria meningitidis and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
XX against Neisseria infections e.g. bacteraemia and meningitis -
XX
PS Claim 11; Fig 26A; 240pp; English.
XX
CC The present invention provides the protein and coding sequences of

```


Db 7269 GCTACGATATCGCGGTGAAGC-----ATTTCTCGCGGTTTT 7228
QY 114 ThrGlyAspMetAlaAspPheSerProAspHisAla-----lleMetVal 128
Db 7227 CAAGCGGACGATCCGATATTATTCGCGAGCGCGTGGGAAAGCGGACAAAGTGGCGCG 7168
QY 129 AspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSer 148
Db 7167 AGTACTGCCAACATCAGGCGGTGGAAATCCCTGAAGGCCCGCTTCCTCGTGTCTTACGGC 7108
QY 149 SerGlyAsnValAlaGlyLeuValAlaAspGlyLysIleProGluLysMetPro 168
Db 7107 CGCAACACGCGCGCGGTGTCATCAGCATGGTTCAGCAATATACGCCAATCTCAACAAAGC 7048
QY 169 GluAsn-----GlyValSerGly----- 174
Db 7047 CGCAACATCGGAGCGGTTTACGGCTCATGGGCAACCGCAGCGCTGAATATGACATTAAC 6988
QY 175 -----GluLeuGlyLeuArgLeuSerSer-----GlyAsnLeuGluLys 187
Db 6987 GAAGTGTGCAACAAACATCCGCCATCCCTCCTCACCAGCGAAGTCGGCGCGCAATTGC 6928
QY 188 LeuThrSerGly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThr 205
Db 6927 TTCCGCGAGCGGATAGACAGCAAAATGTCATGGTTTCGCCAGCATTTACCGTCAAACTC 6868
QY 206 GluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225
Db 6867 GACAACGGCTGAAGTGGAGCGGGCAATACACC-----TAGCACAATGTGGAGCGC 6817
QY 226 LeuProAsp---SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrp 244
Db 6816 ACGCCGACGCGAGTCGC----- 6799
QY 245 ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPro--- 263
Db 6798 ACCAAGTCGCTGAC-----GACCCTTCGAGCTGCGCTTAC 6763
QY 264 -----AlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
Db 6762 CGCATGGGGTTCGCCACCGGAACGATTT----- 6733
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluAsp 299
Db 6732 -----GTCAAGACAAAGTGCAGTTTCG-----CGTTCGCGAC 6700
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAla 319
Db 6699 CTTGAATAC----- 6691
QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db 6690 ---GCCTTCAACGACAAA---TGG-----CGTGC 6667
QY 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db 6666 CAATGGCAG-----CTCGCCCAACCGCAGCGCGCGCGAGAT 6631
QY 360 TyrHisHis-----AspGluLysAlaGlyAspAlaValGluAsnPhePhe----- 374
Db 6630 TTTGATCATTCTTATGACGAGCGCGAAATGGCACTTAATCAACAGTAACTACGCCCTGG 6571
QY 375 -----AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg---HisGlnProIle 390
Db 6570 CAGCAGCCGACACAAACCTGTGCTCCAACTTAACGCTCAACGGCGGACTACACCATC 6511
QY 391 GlyArgLeuLysGly-----SerTrpGlyValGlnTyrLeuGlyGlnLys----- 405
Db 6510 GGCGGTTTTGAATAACCATCGTACCGTACGATGGATTACAGCGCGGAACCGCAACCG 6451
QY 406 -----SerSerAlaLeuSerAlaThr----- 412
|||||

Db 6450 ACATTGGGTTTCAGCAGCGCCTTTCCGCGCTCCATCAACCCCTACAGCCGCGCAACTGG 6391
QY 413 ---SerGluAlaValLysGlnProMetLeuLeuAspAsn-----LysValGlnHisTyr 429
Db 6390 CGGCTTCGGGAGATTGAGCTATTCTGACCCAAACCGCCACAAAGCCGACTCTCTAC 6331
QY 430 SerPhePhe-----GlyValGluGlnAlaAsnTrpAsp---AsnPheThrLeuGluGly 446
Db 6330 GGCATCTTTGTGCAAAACATCTTCCGCGACCGCGATTGTGAATTCGCTCGCGCGC 6271
QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArg 466
Db 6270 -----CGTTCGACAAATACACCTTTAATTC 6244
QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---ThrAlaArg 485
Db 6243 GAAACAAACTC-----ACCGGACGAGCGGCCCAATACAGCGGACAC 6202
QY 486 SerPheAla-----LeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db 6201 TCGTTTCAGCCCAACATCGCGCAGTGTGAACATCAATCCCGTCCAC-----ACACTT 6148
QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db 6147 TACGCTCGTATAACAAA-----GCCTCGCGCCTTATGGCGGACGCGCGCTAT 6097
QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db 6096 TTGACATCATGATACGTTGCTTCCGCGGTTCACGCGCGACCCCGAGTACACCCGCCAA 6037
QY 544 IleGluLeuAlaLeu-----GlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeu 561
Db 6036 TACGAAACCGCGCTGAAAAAGCAGTTGGCTGGACGACCGCTCAGCACTACGTTGTCTGCC 5977
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db 5976 TAC----- 5974
QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
Db 5973 -----CAAAATCGAACGCTTCAAT----- 5956
QY 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArg-----TyrArg 618
Db 5955 -----ATCGCTACCGCGCGGATCCAAAAAACAACCTTATATT 5917
QY 619 IleGlyValSerGlyAsp----- 624
Db 5916 TATCGGTTAGCGGCAACACACCGTTCCGCGCGGTGGAAATGTCGCGCATCGGCGCAATC 5857
QY 625 -----TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
Db 5856 ATCCCCAAAAAATCTATCTCGCGGTTCGTTGGCGGTGATGCAGCGCGAAA----- 5806
QY 640 GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
Db 5805 -----GTCGTTGAAGACAAAGAAATCCCGACCGAGTG 5773
QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaHis 679
Db 5772 -----GGCATCCATTG-----AATAATACAGCAACGTTACCGGCAAC 5734
QY 680 LeuAspTyrTyrArgVal----- 685
Db 5733 CTG---TTTTTCCGTTATACCCCGACCGAAACCTTACGCGGAAATCGCGTAAACCGT 5677
QY 686 -----PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
Db 5676 ACAGGCAACGCTACGCTTACCAAGAAATAAAGAGTACTACGCTTCCAGGCTTT 5617
QY 702 HisMetLeuAsn-----LeuGlyAlaAsnTyrArgArg---AsnThrArgTyrGlyGlu 718
Db 5616 GCCCGAGTTGATGCCTGTGGCTGGACCAATAAATGTTAAGTTACCTTTTGGC--- 5560
|||||


```
Db 302426 CGCATGGGTTGCCACCGAAGATTTT----- 302455
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
Db 302456 -----GTCAAAGACAAGCTGCAAGTTTG-----CGTTCCGAC 302488
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
Db 302489 CTTGATATAC----- 302497
QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db 302498 ---GCCTTCACGACAAA---TGG-----CGTGCC 302521
QY 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db 302522 CAATGGCAG-----CTCGCCACCGCACGCGCGGCGCAGGAT 302557
QY 360 TyrHisHis-----AspGluLysAlaGlyAspAlaValGluAsnPhePhe----- 374
Db 302558 TTTGATCATTTCTATGCAGCGGCAAAATGGCACTTAATCAACGTAACTACGCCCTGG 302617
QY 375 -----AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg---HisGlnProIle 390
Db 302618 CAGCAGACCGACCAACAAACCTCTGCTCCAACCTTAACGCTCAACGCGGACTACACCATC 302677
QY 391 GlyArgLeuLysGly-----SerTrpGlyValGlnTyrLeuGlyGlnLys----- 405
Db 302678 GGCCTGTTTGAACACCAACCTGACGTAGCGATTGATTACAGCCCGCAACCGCAACCCG 302737
QY 406 -----SerSerAlaLeuSerAlaThr----- 412
Db 302738 ACATTGGGTTTCAGCAGCGCTTTTCGCTCCATCAACCCCTACGACCGCGCAAGCTGG 302797
QY 413 ---SerGluAlaValLysGlnProMetLeuLeuAspAsn-----LysValGlnHisTyr 429
Db 302798 CCGGCTTCGGGCGAGATTGCGAGCCTATTCTGACCCAAACCGCCACAAAGCCGACTCCCTAC 302857
QY 430 SerPhePhe-----GlyValGluGlnAlaAsnTrpAsp---AsnPheThrLeuGluGly 446
Db 302858 GGCATCTTTGTGMAAACATCTTCTCGCCACGCCCGGATTTGAAATTCGTCCTCGGCGGC 302917
QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArg 456
Db 302918 -----CGTTACGCAAAATACACCTTTAATTC 302944
QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---ThrAlaArg 485
Db 302945 GAAACACAAACTC-----ACCGCAGCAGCGCCCAATACAGCGGACAC 302986
QY 486 SerPheAla-----LeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db 302987 TCCTTCAGCCCCCAACATCGCGCGAGTGTGAACATCAATCCGTCAC-----ACACTT 303040
QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db 303041 TAGCGCTCGTATAACAAA-----GGCTTCGCGCTTATGCGCGACCGCGGCTAT 303091
QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db 303092 TTGAGCATCATGATTGCTTCGCGCGTTCACGCGCCAGCCCGGAGTACACCCGCCAA 303151
QY 544 IleGluLeuAlaLeu-----GlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeu 561
Db 303152 TACGAAACCGCGGTGAAAGCAGTGTGGCTGGACCGCTCAGCACTAGCTTGTCTGCC 303211
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db 303212 TAC----- 303214
QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
Db 303215 -----CAAATCGAACGCTTCAAT----- 303232
```

```
QY 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArg-----TyrArg 618
Db 303233 -----ATCCGCTACCGCCCGCATCCAAAAACAACCCCTATATT 303271
QY 619 IleGlyValSerGlyAsp----- 624
Db 303272 TATGCGGTTAGCGGCAAAACACCGTTCGCGCGGTGGAAATTGTCGCCATCGGCAAAATC 303331
QY 625 -----TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
Db 303332 ATCCCAAAAAAATCTATCTGCGCGTTCGTTGGCGGTGATGAGCGGCAAA----- 303382
QY 640 GluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
Db 303383 -----GTCGTTGAAGACAAAGAAAAATCCCGACCGAGTG 303415
QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db 303416 -----GGCATCCATTG-----ATAATACCAGCAACGTTACCGGCAAC 303454
QY 680 LeuAspTyrTyrArgVal----- 685
Db 303455 CTG---TTTTCGTTATATCCCGACCGAAACCTCTACCGCGAAATCGGCGTAACCGGT 303511
QY 686 -----PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
Db 303512 ACAGGCAAAACGCTACGCTTACAACTCAAGAAATAAAGAGTACTACGCTTCCAGGCTTT 303571
QY 702 HisMetLeuAsn-----LeuGlyAlaAsnTyrArgArg---AsnThrArgTyrGlyGlu 718
Db 303572 GCCCGAGTTGATGCCATGCTTGCTGGAACCATAAAAATGTTAACGTTACCTTTGCC--- 303628
QY 719 TrpAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSer 738
Db 303629 -----GCAGCCAATCTGCTCAATCAAAATATTGCGCGTTCGCACTCT 303670
QY 739 PheLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 303671 ATG-----CCGGTAATCCGCGCGGTACTACTCCCGGGTAAATTACCGTTTC 303718
```

Search completed: November 17, 2003, 18:17:17

Job time : 3725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 17:11:51 ; Search time 99 Seconds

(without alignments)
3379.479 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTLKPIVLSILLINTEPL.....FLSDTFQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0936377/runat_14112003_104450_22141/app.query.fasta_1.903
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0936377@cgn_1_1_44@runat_14112003_104450_22141 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match %	Length	DB ID	Description
1	1391	34.5	2115	4	US-09-328-352-1018
2	685	17.0	2157	4	US-09-252-991A-12861
3	405.5	10.0	1395	4	US-09-252-991A-13078
4	319.5	7.9	2226	4	US-09-328-352-1772
5	255	6.3	2262	4	US-09-252-991A-1416
6	235.5	5.9	2370	4	US-09-252-991A-10440
7	230	5.7	2169	5	PCT-US96-05320A-264
8	230	5.7	1830121	4	US-09-557-884-1
9	230	5.7	1830121	4	US-09-643-990A-1
10	224.5	5.6	1827	4	US-09-252-991A-1375
11	223	5.5	2381	3	US-08-628-434-3
12	219	5.4	2232	5	PCT-US96-05320A-113

SUMMARIES

Alignment Scores:
Pred. No.: 5.7e-135
Score: 1391.00
Percent Similarity: 56.19%
Best Local Similarity: 40.59%
Query Match: 34.46%
DB: 4

US-09-936-377-2 (1-758) x US-09-328-352-1018 (1-2115)

QY 9 lleValLeuSerIleu---LeuIleAsnThrProLeuLeuAlaGluAlaHisGluThr 27
Db 61 TTGACCTTATCCATTTTGGCTGCCATATCTATCTGTTGTCAGCGAAATGAAAT 120
QY 28 GluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThr 47
Db 121 GTTGAGAAG-----TTAGAACTATCCGATC-----AGGCCGACCCC----- 159

ALIGNMENTS

RESULT 1

US-09-328-352-1018
; Sequence 1018, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Berton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1018
; LENGTH: 2115
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1018

Sequence 12930, A
Sequence 6426, Ap
Sequence 1, Appli
Sequence 1, Appli
Sequence 14350, A
Sequence 14267, A
Sequence 970, App
Sequence 1192, Ap
Sequence 708, App
Sequence 6, Appli
Sequence 194, App
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 14238, A
Sequence 5404, Ap
Sequence 2, Appli
Sequence 1, Appli
Sequence 1816, Ap
Sequence 174, App
Sequence 1708, Ap
Sequence 2016, Ap
Sequence 9244, Ap
Sequence 15877, A
Sequence 14364, A
Sequence 10725, A
Sequence 8394, Ap

US-09-252-991A-12930
US-09-252-991A-6426
US-08-628-434-1
US-08-425-843-1
US-09-252-991A-14350
US-09-252-991A-14267
US-09-221-017B-970
US-09-252-991A-1192
US-09-328-352-708
US-08-425-843-6
US-09-453-702B-194
US-08-487-890A-4
US-08-478-435-4
US-08-337-483-4
US-08-478-373-4
US-08-474-671-4
US-08-483-577A-4
US-08-897-438-4
US-08-637-654-4
US-08-649-518-4
US-09-252-991A-14238
US-09-252-991A-5404
US-09-307-973A-2
US-09-307-973A-1
US-09-328-352-1816
US-03-453-702B-174
US-09-252-991A-1708
US-09-252-991A-2016
US-09-252-991A-9244
US-09-252-991A-15877
US-09-252-991A-14364
US-09-252-991A-10725
US-09-252-991A-8394

c 13 216 5.4 618 4
14 210.5 5.2 2049 4
15 207 5.1 2600 3
16 203 4.061 3
17 198 4.9 1854 4
18 198 4.9 2943 4
19 196 4.9 4661 4
20 193.5 4.8 2391 4
21 193 4.8 2124 4
22 190.5 4.7 4651 3
23 188 4.7 9057 4
24 180.5 4.5 5099 1
25 180.5 4.5 5099 2
26 180.5 4.5 5099 2
27 180.5 4.5 5099 2
28 180.5 4.5 5099 3
29 180.5 4.5 5099 3
30 180.5 4.5 5099 3
31 180.5 4.5 5099 4
32 180.5 4.5 5099 4
33 180 4.5 2271 4
34 180 4.5 2403 4
35 176 4.4 1844 4
36 175 4.3 1845 4
37 174 4.3 2286 4
38 174 4.3 7304 4
39 173.5 4.3 2172 4
40 173.5 4.3 2241 4
41 173 4.3 1788 4
42 170.5 4.2 2121 4
43 169.5 4.2 3738 4
44 169 4.2 1953 4
45 168.5 4.2 3345 4

Db 375 ATCCGGACCTCGAAGCCCGTTTCACTCGCTGAGCTGAGCCGCCCTATACCAAGTAC 316
 Qy 361 HisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsn 380
 Db 315 GAGCACAAGAAATCGAGGATGCGGAGACCGGACACCTTCAAGAACGAGGCTACGAA 256
 Qy 381 AlaArgIleGluLeuArgHisGlnProIleGlyArgGluLysGlySerTrpGlyValGln 400
 Db 255 GGCGCATCGAGCGCCGACCGCCGCTCGGCGCTGAAACGGGTGGTGGCGCGCGAG 196
 Qy 401 TyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet 420
 Db 195 TTCGCCAACGCCCTCTTCGCGCCCTCGG-----GAGGAGCGCTTC 154
 Qy 421 LeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp--- 439
 Db 153 GTGCCGACACGGAACCGACGAGCGCGCTGTTCCGCCCTGGAGAA-----TGGAAAG 100
 Qy 440 -----AspAsnPheThrLeuGluGlyGlyValArgValGlu 451
 Db 99 CTCAGCAGCCGCTCGACCTCAGCTTCGCGCGCGCGCTGGAG 58
 RESULT 4
 US-09-328-352-1772
 ; Sequence 1772, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 1772
 ; LENGTH: 2226
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-1772
 Alignment Scores:
 Pred. No.: 3,68e-23 Length: 2226
 Score: 319.50 Matches: 164
 Percent Similarity: 36.38% Conservative: 127
 Best Local Similarity: 20.50% Mismatches: 333
 Query Match: 7.92% Indels: 176
 DB: 4 Gaps: 30
 US-09-936-377-2 (1-758) x US-09-328-352-1772 (1-2226)
 Qy 59 LysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAsp 78
 Db 46 AAGTAGAAGTCCCAATACA-----ATAGGTGATGCTCTTAAA 84
 Qy 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArg 98
 Db 85 AATATAACAGGTATACAAAGTACTTTCATTTCGACCCCAATGCGAGTGCACCGATATACGT 144
 Qy 99 GlyThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAla 118
 Db 145 AGTTTGAGTGAATTCGGGTGGAGTGAATGAAATGAGAAATTAATTAATGGAATGAAT 204
 Qy 119 AspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIle 138
 Db 205 GCATTAGTGACATATTAATATACCATTCGATCCCAATTTTATAGAAAAGTATAGTAGT 264
 Qy 139 LeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal 158
 Db 265 AACAAAAATACAGATAATATTCGTTATGGCGGAAATCAATAGGCGGAGTGTCAAATA 324
 Qy 159 AlaaspGlyLysIleProGluLysMetProGluAsn----- 170
 Db 325 GAATCGGTTTAAATTCCTTAAATAATTGAAGAAATTAAGTATATGTCCTTT 384

Qy 171 -----GlyValSerGly----- 174
 Db 385 CGGAAAGGATTTAATGATTTTGAAGGGTTCATTTTAAACATCAATGATCAAAAA 444
 Qy 175 -----GluLeuGlyLeuArgLeuSerSer-----Gly 183
 Db 445 AACTGGTCTCAAAATATAAGATATTTCAGAAATATGAATTTCTTTATAAAATCCAGGA 504
 Qy 184 AsnLeuGlu---LysLeu-----ThrSerGlyGlyIleAsnIle 195
 Db 505 AATAGTAAGCCCAANTATGTGAAGCAGAGATTTTTCAAATTCAGGTGGGATCAATAGT 564
 Qy 196 GlyLeuGlyLysAsnPhe-----ValLeuHisThrGluGlyLeuTyrArg 210
 Db 565 GCATTAGCAGCTTCTGTCAAAAGATAGTAGTCCCAACAT-----ATTATATAC 615
 Qy 211 LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro 230
 Db 616 AAGTCGTCAACACCATACATAGATAAATTT-----ATGACTGAGAATCCCTGAT----- 663
 Qy 231 ArgArgPheAlaAsnGlyGln-----HisArgAlaValLeuGlyTrp--A 245
 Db 664 -----TGGGCAGATGGCGATTTTCTTTTATACAGATAAGCCACATCTATATGGGA 717
 Qy 245 ArgArgPheTyrArgArgThrTyrSer-----A 255
 Db 718 GGAATAACATATATAAATCCAAAAATCCAGAGTACATACCTAATCTCCGCAAAATACA 777
 Qy 255 spArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla 275
 Db 778 ATAAAAAGATCAATACGGA-----TGTACACCA 807
 Qy 275 spIle-----IleTrpGlnL 280
 Db 808 AATATTATTTAAAAAATTAGGTAAACAGTTATGCTCAGAAATGAGAATATTGGGTTCGAACA 867
 Qy 280 ysSerLeu--IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
 Db 868 ACATACTTTTGTATAAAGTTTCATGGATTG-----AGTGCAGATAGAAAAA 918
 Qy 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
 Db 919 AGTGAATATGGTGTCACAGATTTTCATTA-----CAAAATCAATCTTTTGTCT 966
 Qy 320 HisAlaHisAsnGlyLysPro---TrpIleAspLeuArgAsnLysArgTyrGluLeuArg 338
 Db 967 GATTCTTATGAACAAATACCGGTGGTGTGAAAAATAGATCAAAATCGTTTGTATTAAAT 1026
 Qy 339 AlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn 358
 Db 1027 TCCAAATTTTATCCAACTATTCCTCTGCAGAAAGAGATAAGTTTGAATTTTCAACAGCTT 1086
 Qy 359 AspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThr 378
 Db 1087 TCTAATAAATATCGGAGATATCTGTGACTGCAAAAGCTTAATAGATATAAATTCATAAT 1146
 Qy 379 GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGly 398
 Db 1147 CATTAATTAATTAATTAATGAACAGTCACTCTTTAAGGGATTAGATCGGATCTGTGT 1206
 Qy 399 ValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGln 418
 Db 1207 TTTAGC---CTTAAAAATAGGAATATCGAAGGAGTGGTACTCAGCGCTATTTCCTAAT 1263
 Qy 419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
 Db 1264 GTAAGTACTATTAGTAAGACTATA-----TTCTTACAGAAGAAATTAAT 1308
 Qy 439 TrpAspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArg 458
 Db 1309 ATCAACAATTTACTTTAAATACGGGCTATCGTTTGAAGAAATTAAGATCATGAGCTT--- 1365

Db 748 AAGGGCGCGACTAC----- 762
 Qy 231 ArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgPheTyrArg 250
 Db 763 ---CGGACGGCAACACGACGATATCGACGAGTCTCTCAGACCCATTGGGAG 819
 Qy 251 ArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp 270
 Db 820 CTCACCGACAGCGACAG-----TTGGCGGCCAACTTCCACTACTAC 861
 Qy 271 AspCysHisAlaAspIleLeuTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeu 290
 Db 862 GACGCTACCGCGATATG----- 879
 Qy 291 TyrProHisLeuLeuThrGluAspValAspTyrAspAsnProGlyLeuSerCysGly 310
 Db 880 ---CCGGCGGCTGACCCAGCGCGAGTACGAC---GACGATCCCTTCCAGTCGTA--- 930
 Qy 311 PheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeu 330
 Db 931 -----CGCGACTGGGACCAATTTC 948
 Qy 331 ArgAsnLysArg----- 334
 Db 949 CTGGTGGCGGCAAGGACTTCTCGTGAAGTACACCGCGAGTCCGACGACCTCACCAG 1008
 Qy 335 TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
 Db 1009 TTCGAGGTGCTCACTACTACAGCAGCAGTTCCTCGCGGCGAGCAGCATCCCGCGCGCAAC 1068
 Qy 355 LeuAsnArg-----AsnAspTyrHisAspGlnLysAlaGlyAspAla 369
 Db 1069 CTAGGACCATACCTCGTACCGCGGACTACATGTGTTCGCGTGAGCGCGCGTTC 1128
 Qy 370 ValGluAsnPhePheAsnAsnGlnTrpGlnAsnAlaArgIleGluLeuArgHisGlnPro 389
 Db 1129 TCGCGGATCTTCTTCGCGCGCCGACCC-----CAGGAG 1164
 Qy 390 IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
 Db 1165 GTCGC-----ATCGGTACCGTACTTCTGAAGAA----- 1194
 Qy 410 SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn----- 424
 Db 1195 ---GCGATGACGAGCGCGCGCAGCTGCGCCTGTGTGACATGTGCGGACCGTGCCT 1251
 Qy 425 -----LysValGlnHisTyrSerPhe 431
 Db 1252 CCGGCTCGGACGCCCATACCTACAGACCGTACCGCGGTACCGAGGCCAGCGCTTC 1311
 Qy 432 PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGlyGlyValArgValGlu 451
 Db 1312 TACATGACGACAGATCGATGTCGGGACTGACGACCATACCCCGGCGATCCGCTTCGAG 1371
 Qy 452 LysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle----- 464
 Db 1372 AAGATGACAGCGACTGGCGC---GACCGCGCGTGTCTCGGCTCGGCAACGCGCGGTG 1428
 Qy 465 -----AspArgGluAsnTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481
 Db 1429 CAGGAGAGAGCGCAGCAGGACTACACGACCGCTGCGCGCTG----- 1476
 Qy 482 GlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu 501
 Db 1477 -----AGCGTGATGTATCACTTACGCGCAGGTGG----- 1506
 Qy 502 SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGly 521
 Db 1507 AAGCTGTTCGCCCACTACTCGCAATCGTTCGGCAGCGCTGAG---TACTTCCAGCTCGGC 1563
 Qy 522 LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
 Db 1564 CAGGCGCGCAGCGCACGATACCGCGCGCGC-----CTGGAGCGCGAGAGGCC 1614

Qy 542 AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeu 561
 Db 1615 AAGACCTACGAAGTCTGGCGACGCTACGACACGCGAGTGGGCGCGGAGATCACGCTG 1674
 Qy 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
 Db 1675 TTC----- 1677
 Qy 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
 Db 1678 ---TACATCGACTTCGACGAGAGTCTCAGTACGTAGC---AACGACGTCGGCTGAC 1731
 Qy 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyVal 621
 Db 1732 AACCTCGCGCGC-----ACCAAGCACCAGGCGCTCAGACT 1767
 Qy 622 SerGlyAspTyr-----ValArgGly 628
 Db 1768 TCCGGCAGTATGACTTCGCGCTCTCGATCCGCGCTGGACGCGCTGCTACGCG 1827
 Qy 629 ArgLeuLys-----AsnLeuProSerLeuProGlyArgGlu 640
 Db 1828 AGCCTACCTATACCGCGCCACCTACGAGCGCAGCATTCCTCTTCAAGGGCGGAC 1887
 Qy 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
 Db 1888 CTGCGCTCTACTCGCGC-----CAGTGGCC 1914
 Qy 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
 Db 1915 ACCGCGGGTGGCTACGAGTGC-----GACCTGTGACCTACACCTC 1959
 Qy 681 AspTyrTyrArgValPheAlaGlnAsn-----LysLeuAlaArgTyrGluThrArgThr 698
 Db 1960 GAC-----GCCTTCGCGCAGTCCATGACGCGCGC-GCAGGGCTGACGACCGCAG 2009
 Qy 699 ProGly-----HisMetLeuAsnLeuGlyAlaAsnTyrArgArg 712
 Db 2010 CCAGGCAACTTCCACCACTACATCCGAGCCCGCGCGCCGCGCTACGCGCA 2069
 Qy 713 AsnThrArg-TyrGlyGluTrpAsnTrpTyrValLysAlaAsp----- 726
 Db 2070 CATTCGCGCTAGCGTACCTGGAATGCCGCTCGGTACGACTTCGGCCCGCAGCGCTC 2129
 Qy 727 -----AsnLeuAsnGlnSerValTyrAlaHis 737
 Db 2130 GAACCTGAAGTCTGCGGCTGAGAACCTGTTCGACGAGCAGTATTTACCCCGCTC 2189
 Qy 737 rSer 738
 Db 2190 CAGC 2193
 RESULT 6
 US-09-252-991A-10440
 ; Sequence 10440, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10440
 ; LENGTH: 2370
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10440

Alignment Scores:

Pred. No.: 9,03e-15 Length: 2370
 Score: 239,50 Matches: 188
 Percent Similarity: 34.23% Conservative: 118
 Best Local Similarity: 21.03% Mismatches: 294
 Query Match: 5.93% Indels: 295
 DB: 4 Gaps: 42

US-09-936-377-2 (1-758) x US-09-252-991A-10440 (1-2370)

QY 6 LeuLysProIleValLeuSerIleLeuLeuLeuAsn----- 17
 Db 100 CTGGGCCCC--TGCTGGCCCTGCTGTGCTCAGCCCTTCCCTGGCCCTGGCGGGGAAC 156
 QY 18 -----ThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeu 33
 Db 157 GCATCCCGCTGACCCCGACCCACCACATCACCACCACCGTACCGAGGAGGAGTGGATTGG 216
 QY 34 GluThrValThrValValGlyLysSerArgProAlaAlaThrSerGlyLeuLeuHisThr 53
 Db 217 GTGCCAAGCACCGCTCAGCTGCGGACCGCGAAACAACTGGAC----- 258
 QY 54 SerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeu 73
 Db 259 -----CGGCAGAACGTCACAAACATC 279
 QY 74 GlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnIleGlyGlyAlaSer 93
 Db 280 AAGAACTGTTGGCTACGAACCGGAGTC-----TCGGTCGGCGCGCCGACGCT 333
 QY 94 AlaProVal-----IleArgGlyGlnThrGlyArgArgIleLysValLeuAsn 109
 Db 334 GCGGGATCAGCGGTACAAACATCCCGCGCATCGACGGGAACCGCATCTTACGCAGATC 393
 QY 110 HisHisGlyGluThrGlyAsp-----MetAlaAspPheSerProAspHisAlaIle 126
 Db 394 GACGGGTGCACTGCCACGACTTCTTCAGCGGCGCCCTACGCGCAGACCCACCGCAAC 453
 QY 127 MetValAspThrAlaLeuSerGlnGlnValGluLeuArgGlyProValThrLeuLeu 146
 Db 454 TAGCTCGATCCGACATCGATAAAGCGGTGGAATCTTTCGGCGCGCGCTCGGCGCTA 513
 QY 147 TyrSerSerGlyAsnValAlaGlyLeuValAsp----- 157
 Db 514 TAGCGCAGCAACGCCATCGCGCGCGCGGTGAGTACTTCACCTCGACCCGCTCGGACATC 573
 QY 158 ValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGly 177
 Db 574 ATCAAGAGCGCAAG-----GACGTGGC 597
 QY 178 LeuArgLeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyIle 193
 Db 598 GCGCGCTGAAGCGCGGCTACGAGTGGCCAGCCACTCTCGGTGACCTCGGCCACCGTC 657
 QY 194 AsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu-----Tyr 209
 Db 658 ---GCGCGCGCGCGACACTTC-----GACGGCTGTGCTATTATGGCTAC 702
 QY 210 ArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSer 229
 Db 703 CGCCAGG-----CCACGAGACCGAATCCACCGCGG----- 734
 QY 230 ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyr 249
 Db 735 CCACGCGG-----CACCGGGCTCTCGCG----- 758
 QY 250 ArgArgThrTyrSerAspArgArg-AspGlnTyrGlyLeuProAlaHisSerHisGluTyr 269
 Db 759 CAGCGAGCAACCGGAGACCGCGACAGCTACAGCCTG----- 798
 QY 269 rAspAspCysHisAlaAspIleIleTrpGlnLysSer-----Leu11 283

Db 799 -----CTGGCAAGCTGGCTGGAATACGCCGAGGCTAGCGCTTCGGGTGT 848
 QY 283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
 Db 849 CTTTGAGAAATAC-----AAGAGCAGCTCGATACCGA 881
 QY 303 pAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAs 323
 Db 882 CCAGAAGAGCGCTATGGCGGCCGTACGAC-----AA 914
 QY 323 nGlyLysProTrpIle----- 328
 Db 915 GGCAGAGCGCGCATCCCGCAGCATGTCGCGCGGCATGTACAGTGGCGCAAGGG 974
 QY 329 -----AspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPhePr 346
 Db 975 CAAAGCACCCCTGACTCGCGAGCGCTACGCGCTGGAGCCATTTCTGCTCGACAGCCC 1034
 QY 346 oGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAl 366
 Db 1035 GGTGCGCGATCGATCCAGTGGAGCCTGAAC-----TACCAGTTGGCGAGAC 1082
 QY 366 aGlyAspAlaValGluAsnPhe----- 373
 Db 1083 CGACCGAGCGACCGCGAGTTCTTACTACCGCATCACCAGGCTCTCGCACCGCGA 1142
 QY 374 -----PheAsnAsnGlnThrGlnAsnAla----- 381
 Db 1143 CACTACTACAGGAACGCTGTGGGTCTTCAGACCGAGTTGGACAAGAGCTTCGCCAT 1202
 QY 382 -----ArgIleGluLeuArgHisGlnProIleGlyAr 392
 Db 1203 CGCGAGACCGAGCACCTGCTGAGCTACGGGATCAATCTCAAGCACCAAGAGTCCACCG 1262
 QY 392 gLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaTh 412
 Db 1263 CATGCGACGCGACCGGCCCAATCTGGACACCGCGCGAGCAGTCCGCGCATCCCT 1322
 QY 412 rSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePh 432
 Db 1323 GGAGCGCAGCAGCTTTCCC-----GATCGAGGTGAAGACTTACGCTGT 1373
 QY 432 eGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGlu-- 451
 Db 1374 CGCCAGGACAGCATCAGCTGGAACGACTGACCTTCACTCCGCGCTCGTTACGACTA 1433
 QY 452 -----LysGlnLysAlaSe 456
 Db 1434 CACGCGATGAGCGCGACATCACCAGAGTTCCTGCGCACCATGAAGCAGACGAGAA 1493
 QY 456 rIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAs 476
 Db 1494 CACCGCGTTCAGAGTCG-----GACAAGAATGGCACCAGCGGTTCGCCAA 1541
 QY 476 pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheTh 496
 Db 1542 GTTCGCG-----GTGACCTACGACTTCGCCAGCAGCTACACCTGGTAC----- 1584
 QY 496 rProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGl 516
 Db 1585 -GGCCAATACGCCCGAGGCTTC-----CGCACGCCCGCCAGGAGC 1625
 QY 516 uLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLe 536
 Db 1626 GCTGTACGCTGATTCGAGAACCTTCAGCGCGGCTACCATCGAGCTTACCCCACT 1685
 QY 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGl 556
 Db 1686 CAAGCCGAAAAGACCCAGAGCTTCGAGACCGGGCTCGCGCGAGTTTCGAGGAGGAGCAG 1745
 QY 556 nTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAs 576

QY 184 AsnLeuGluLys-----LeuThrSer 190
 Db 517 TTATCGGAAAGGATGTTCTGTATTTGCGCAAAATGACAAATTCGATGTTCTTTATTAGT 576
 QY 191 Gly-----GlyIleAsnIleGlyLeuLysAsnPheValLeuHisThrGlu 206
 Db 577 GGTTCCTATATATGCGGATAATTTACGCACTGGTAAAGGCAACAGCTAAATATACC 636
 QY 207 GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg-----Asn 222
 Db 637 GCC---TATAAACAGTTTGGGGCTTAGCAAAATTCGGTTGCGCAAAATTAATGATCGCAAC 693
 QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaHisGlnHisArgAlaValLeu 242
 Db 694 CGCGTGAATATATCCACCGCAAACTCGTTTAA---AACAAACAGACCAAGCAATATAG 751
 QY 243 GlyTyrArgLys-----ArgPheTyrArgTyrArgTyrSerAsp 255
 Db 752 AGGTGGAACGAACTTACTAATGACAAATTCAGATCAAAATCAAAAGTTCCACGGAC 811
 QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp----- 271
 Db 812 AAAAAGACGATCTTCTCTCTCAACACACCAACCATCACCATCAGAAAGATCAGAGTTT 871
 QY 272 -----CysHisAlaAspIleIle-----TrpGlnLysSerLeu 282
 Db 872 ACTCTAAGTGAACACGTTTAGTGTAGTGTACGTATTTAACTGATCAACAAATTCCTG 931
 QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
 Db 932 ATCAAGCACGGTATTAAAC-TATTAT-----TTAAGCCA----- 966
 QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
 Db 967 GATAATCCTTATCTA-----AATACGCATATCGCACTGTAT 1002
 QY 323 AsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
 Db 1003 AACATAAACTATTGAGAAGACAGCGTAAAGTGTGTGTGTTGAAAGATCAGACTAAA 1062
 QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
 Db 1063 TTGACTACCGGAGT-----ATAAATTACGTAATCTCTCGAATTATCT 1107
 QY 359 -----AspTyrHisHisAspGluLysAlaGlyAspAlaVal 370
 Db 1108 CACATTTCTTGTGTATGGGTGGATATATATCGAGATATAAATCCGTACCGAA----- 1161
 QY 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProile 390
 Db 1162 -----CGAGGCACAAACGGTAGCGATCGAGTTTCGAGCGGACCCCTAT 1206
 QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSer 410
 Db 1207 AATGCGAATTCAACACTACAGCGGTT---TATTAA----- 1239
 QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
 Db 1240 -----ATCGCCCATATTCG 1254
 QY 431 PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgVal 450
 Db 1255 CTATTTGG----- 1263
 QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
 Db 1264 GAAAAATTTGCTAGTTTCGCCAAGGTACGTATGACCACCTACGTACCTCAAGTAAAC 1323
 QY 463 LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
 Db 1324 GTAAATACAGGATAATCAITTA----- 1347

QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTyrPyrPheThrProGlnHisLysLeuSer 502
 Db 1348 -----TCTCTGCCCAAAAATTAACCTGGATAGTGACC-----AATGGTTAGAT 1392
 QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 Db 1393 TTTACTGCCAAATATAATAGAGCTTTCCGAGCACCATCTATGCAAGACGCAATTTGTGAGT 1452
 QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
 Db 1453 GGTGCTCACITTTGGGGCAATATCTCTAGGGCTAGATCACATCATAGATTTGTAGCAAT 1512
 QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
 Db 1513 CCAAAATTTGGCCCTGAAACAGCGAAATAAAGAAATTAACCGCAATCTACATTTTGTAT 1572
 QY 547 AlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
 Db 1573 AGTCTGTTTAAACAGGCGATATAATCAAAATTTGAAGCGCATTTATTTCCGTAATGATGTG 1632
 QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
 Db 1633 AAAGATTTTATTAACCTTAAATAATTTATGATGCAAGACAGTGCAGTGCAGGTGCA 1692
 QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595
 Db 1693 AATCCAAATACAAATGAGGACATTTGTGCCAAAAAATTCAGATATCAA----- 1740
 QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
 Db 1741 AACATAACTAATGCCCGTTTAAGCGGTATTGAATTGCAAGCTCAATACCAA---ACAGAA 1797
 QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
 Db 1798 CGTTTAAACGCTA-----TTTACTAATCTATGGCAGC 1827
 QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
 Db 1828 ACCAAAGGTAAAGATAAAGATAGTGGC-----GAAAGCT 1860
 QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
 Db 1861 TTATCAACATTTGCCGACAGCAAAATCGCGTAGGGTAAATTTATGCTTTAGTAAAAAGAC 1920
 QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
 Db 1921 AAATTCACGTGGGAGCGACAGTAAACCCATTACGCT-----GCTCAACGCCGAGTGCCT 1974
 QY 693 ArgTyrGluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
 Db 1975 AAAGATCATAGTGTATCTACCCCAAGTTATATATAGTACCGCATCTTCGTCTACCTATGCA 2034
 QY 712 ArgAsnThrArgTyrGlyGluTyr 719
 Db 2035 -----CCATTAAAGCGCAATGG 2052

RESULT 8

US-09-557-884-1
 ; Sequence 1, Application US/09557884
 ; Patent No. 6505581
 ; GENERAL INFORMATION:
 ; APPLICANT: Fleischmann et al.
 ; TITLE OF INVENTION: The Nucleotide sequence of
 ; the Haemophilus influenzae Rd Genome, Fragments
 ; Thereof, and Uses Thereof
 ;
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: MD
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS v6.22
 SOFTWARE: ASCII Text
 CURRENT APPLICATION NUMBER: US/09/557,884
 FILING DATE: 25-Apr-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/476,102
 FILING DATE: JUN-5-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Michelle S. Marks
 REGISTRATION NUMBER: 41,971
 REFERENCE/DOCKET NUMBER: PB186P3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-309-8504
 TELEFAX: 301-309-8439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1830121 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 US-09-557-884-1

Alignment Scores:
 Pred. No.: 4,79e-09 Length: 1830121
 Score: 230.00 Matches: 171
 Percent Similarity: 36.63% Conservative: 125
 Best Local Similarity: 21.16% Mismatches: 285
 Query Match: 5.70% Indels: 229
 DB: 4 Gaps: 41

US-09-936-377-2 (1-758) x US-09-557-884-1 (1-1830121)

```

QY 184 AsnLeuGluLys-----LeuThrSer 190
D 292200 TTATCGGAAAGGATGTTTCTGTATTTGCGCAAAATGACAAATTCGATGTTCTTATTAGT 292259
QY 191 Gly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGlu 206
D 292260 GGTTCCTATAATAATGCGGATAAATTTACGCACTGGTAAAGGCAACACAGCTAAATAATACC 292319
QY 207 GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg-----Asn 222
D 292320 GCC---TATAACACAGTTTGGGGCTTAGCAAAATTCGGTTGGCAAAATTAATGATCGGAC 292376
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
D 292377 CCGGTGGAATTTATCCACCGGAAACTCGTTTAA--AAACAACACAGCACAAGCAATAATG 292434
QY 243 GlyTyrArgLys-----ArgPheTyrArgArgThrTyrSerAsp 255
D 292435 AGGTGGAACGAACTTACTAATGAACAAATTAACGATCAATCAAAAGTTCCACGGAC 292494
QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAsp----- 271
D 292495 AAAAGACGATCTCTCTCTCTACACACACACCATCACCATCAGAAAGATCAGAGTTT 292554
QY 272 -----CysHisAlaAspIlele-----TrpGlnLysSerLeu 282
D 292555 ACTCTAAAGTGAACACGCTTTAGGTAGTGTGCTAGTTATTAACTGATCAACAAATTCCTG 292614
QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
D 292615 ATCAAGACAGGTATTAACT-TATTAT-----TTACGCCA----- 292649
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
D 292650 GATAATCTCTTATCTA-----AATACGCATATCGCACTGAT 292685
QY 323 AsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
D 292686 AACATATAAACTATTGAGAAAGACAGCGTAAGTCAGTGGTGAAGATCAGACTAAA 292745
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
D 292746 TTGACTACCCGAGGT-----ATAAATTACGTAATTCCTCCGAATTATCT 292790
QY 359 -----AspTyrHisHisAspGluLysAlaGlyAspAlaVal 370
D 292791 CACATTTCCCTTTGTTTATGGGTGATTTATATGCGAGATATAAATCCGTACCGAA----- 292844
QY 371 GluAsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle 390
D 292845 -----CGAGGCACAAACGGTAGCGATCGCAAGTTTCGAGCGACCCCTAT 292889
QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSer 410
D 292890 AATGCGAATTCACCACTACAGCGTT--TATTTA----- 292922
QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
D 292923 -----ATCCCATATATCCG 292937
QY 431 PhePheGlyValGluGlnAlaAsnTrpAsnPheThrLeuGluGlyValArgVal 450
D 292938 CTATTTGGG----- 292946
QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
D 292947 GAAAAATTCCTAGTTTCGCAAGAGTGTACGTTATGACCACCTACGATACCTCAAGTAAACC 293006
QY 463 LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
D 293007 GTAAATAACAGGATAATCATTTA----- 293030

```

```
QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSer 502
Db 293031 -----TCTCCTGCCAATAATTAACCTGGATAGTACC-----AATGGTTAGAT 293075
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 293076 TTTACTGCCAATATATGAAGCTTTCCGAGCACCCTATGCAAGACGATTTGTGAGT 293135
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
Db 293136 GGTGCTCCTTTGGGCAATACTCTAGGGCTAGATCACATCAATAGATTTGTAGCAAT 293195
QY 534 LysHisLeuAsnLysGlyArgSerAsnAsnIleGluLeu----- 546
Db 293196 CCAATTTGGCCCTGAAACAGCGAATAAATAAGAAATTTACCGCAATCTACATTTTGAT 293255
QY 547 AlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
Db 293256 AGTCTGTTAAACAAGCGGATAAATTCAAAATTGAAGCGACTTATTTCCGTAATGATGTG 293315
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
Db 293316 AAGATTTTATTAATCTTAAATAATTTAATGATGCAAAAGCAAGTGCAGGTGCA 293375
QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595
Db 293376 AATCCAAATACAATGAGCATTTGTCGCAAAAATTTCCAGATCA----- 293423
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
Db 293424 AACATAACTAATCCCGTTTAAGCGGTATTGAATTGCAAGTCAATACCAA---ACAGAA 293480
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 293481 CGTTTAACGCTA-----TTTACTAACTATGCGCAGC 293510
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 293511 ACCAAAGGTAAGATAAAGATAGTGC----- 293543
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
Db 293544 TTATCAACATTTCCGCAAGCAAAATCGCGGTAGGTAATATGCTTTAGTAAAGAC 293603
QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 293604 AATTCAAGTGGGAGCGACAGTAACCCATTACGCT-----GCTCAACCGCGAGTGCCT 293657
QY 693 ArgTyrGluThrArgThr-----ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 293658 AAGATCATAGTGTACCTACCCCAAGTTATATATCTGACCATCTTCTGCTACCTATGCA 293717
QY 712 ArgAsnThrArgTyrGlyGluTyr 719
Db 293718 -----CCATTAAAGCGCAATGG 293735
```

RESULT 9

US-09-643-990A-1

; Sequence 1, Application US/09643990A

; Patent No. 6528289

GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann

; Mark D. Adams

; Owen White

; Hamilton O. Smith

; J. Craig Venter

; TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

QY 147 TyrSerSerGlyAsnValAlaGlyLeuValAlaAspValAlaAspGlyLysIleProGluLys 166

```
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
```

Alignment Scores:

Pred. No.:	4,79e-09	Length:	1830121
Score:	230.00	Matches:	171
Percent Similarity:	36.63%	Conservative:	125
Best Local Similarity:	21.16%	Mismatches:	285
Query Match:	5.70%	Indels:	229
DB:	4	Gaps:	41

US-09-936-377-2 (1-758) x US-09-643-990A-1 (1-1830121)

QY	11	LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer	30
Db	291699	CTTCCCTTCGCAATTACACACC-----TAGTCACAGCAATGCGTAGCGCAATCC	291752
QY	31	ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu	50
Db	291753	GTGTAATTAGACTTATCAACGTTATTGCGACACGAGAYCCA-----AGTAGGTTT	291803
QY	51	LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla	70
Db	291804	GCTTATACCCGAGAAAAACAATCTAAA-----GATAGTCTTCTTCTTAAGCAAGCG	291854
QY	71	ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGly	90
Db	291855	ACTAGTGTTCAGATGCGTTAGAGACATTCCTCCATGTTGATTTAGT-----GCGCGT	291908
QY	91	Gly-----AlaSerAlaProValIleArgGlyGlnThrGlyArgArg-----IleLys	106
Db	291909	TCGAGAAGCATTTGCTCAAAAACCTAATATTCGAGGGTTAAGTGATAATCGTGTGTGCAA	291969
QY	107	ValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle	126
Db	291969	GTCATGTGCGGTGAGCAAAAATTTGATTATTAGCACATAGAGGT-----TCTTAT	292019
QY	127	MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu	146
Db	292020	TTTCTTCCAAATGCTACTCATCCAGAAATTGAATCAAGAGCAAGTAGTAGCTCCTTA	292079
QY	147	TyrSerSerGlyAsnValAlaGlyLeuValAlaAspValAlaAspGlyLysIleProGluLys	166

D	b		:::		:::		:::		:::	TGGGTACGGTGCATTTGGTGCTGTTGGCAATCGGTACGCCAAATGCTTTAGACTTA	292139
Q	y	167	MetProGluAsn-	-	GlyValSerGlyLeuLeuGlyLeuArgLeuSerSerGly	183	:::		:::		
D	b	292140	TTGAATAAATAATACAAAATTTCGGAGTTAAAAATTCGCAAGTTATCAAACGCTAATAAT	292199							
Q	y	184	AsnLeuGluLys-	-						-LeuThrSer	190
D	b	292200	TTATCGGAAGAAGATGTTTCTGTATTTGGCGCAATACAAAATTCGATGTTCTTATTAGT	292259							
Q	y	191	Gly-	-	GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGlu	206					
D	b	292260	GGTTTCTATAATAATGCGGATAATTACCACTGGTAARAGGCAACAGACTAATAATATACC	292319							
Q	y	207	GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg	222						-Asn	222
D	b	292320	GCC--TATAAACAGTTGGGGGCTTAGCAAAATTCGGTTGGCAATTAATGATGCGAAC	292376							
Q	y	223	LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyClnHisArgAlaValLeu	242							
D	b	292377	CGCTGGGAATTATCCCACCGGAACTGTTTTA--AACAAACAGCACCAAGCAATAAATG	292434							
Q	y	243	GlyTrpArgLys-	-	ArgPheTyrArgArgThrTyrSerAsp	255					
D	b	292435	AGGTGAAAACGAACCTTACTAATGAACAATAATACAGATCAAAATCAAAAGTTCACGGAC	292494							
Q	y	256	ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAsp	271							
D	b	292495	AAAAAGACGATCTTCTCTCTCTACACAAACCATCACCATCAGAAAGATCAGAGT	292554							
Q	y	272	--CysHisAlaAspIleIle	-	TrpGlnLysSerLeu	282					
D	b	292555	ACTCTAAAGTGAACACACGTTTAGCTAGTGCAGTTATTAACTGATCAACAAATTCCTG	292614							
Q	y	283	IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValaspTyr	302							
D	b	292615	ATCAAAAGCACGATATTTAAC-TATTAT--TTAACCCA--	292649							
Q	y	303	AspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHis	322							
D	b	292650	GATAATCCTTATCTA-----AATACGCATATCCACTGTAT	292685							
Q	y	323	AsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLys	342							
D	b	292686	AACAATAAAACATTATGAGAAGAACAGAGTAAAGTCAGTGGTGTGAAGATCAGACTAA	292745							
Q	y	343	GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn-	358							
D	b	292746	TTGACTACCCGAGT-----ATPAATTTACGTAATCTTCGGAATATCT	292790							
Q	y	359	--AspTyrHisHisAspGluLysAlaGlyAspAlaVal	370							
D	b	292791	CACATTTCTTTTATGCGGTGATATATATGCGAGATAAAATCCGTACCGAA----	292844							
Q	y	371	GluAsnPhePheAsnAsnGlnThrClnAsnAlaArgIleGluLeuArgHisClnProIle	390							
D	b	292845	---CGAGGCACAAACCGTAGCGATGCGAAGTTTCGACGCGACCCCTAT	292889							
Q	y	391	GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSer	410							
D	b	292890	AATGCGAATTCAACACTACAGCGCT--TATTTA-----	292922							
Q	y	411	AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer	430							
D	b	292923	-----ATCGCCCATATTCG	292937							
Q	y	431	PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgVal	450							
D	b	292938	CTAATTTGGG-----	292946							
Q	y	451	GluLys-----GlnLysAlaSerIleArgTyrAsp	462							

292947 GAAAAATTGCTAGTTTCCCAAGTGTACGTTATGACCACTACGATACCTCAAGTAAAACC 293000
QY 463 LeuIleAspArgGluAsnTyrTyrlsGlnProLeuProAspLeuGlyAlaHisargGln 482
Db 293007 GTAAATACAGGATAATCATTTA----- 293030
QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
Db 293031 -----TCTCCTGCCACAAAATAACTTGGATAGTACC-----AATTGGTTAGAT 293075
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 293076 TTTACTGCGCAATAATATGAAGCTTTCCGAGCACCATTCTATGCAAGAGCGAATTTCTGAGT 293135
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
Db 293136 GGTGCTCACTTTGGGCGAAATACTCTAGGGTAGATCATCAATAGATTGTGTGCAAAAT 293195
QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
Db 293196 CCAAAATTTGCGCCCTGAAACAGCGAAATAAAGAAATTCGCAAAATCTACATTTTGTAT 293255
QY 547 AlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
Db 293256 AGTCTGTTTAAACAGGCGATAAATTCAAAATTTGAAGCGACTATTTCGTAATGATGTG 293315
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
Db 293316 AAAGATTTTAACTTAAATAATTTAATGATGCAAAAGCAAGTGCAAGTGCGAGTGCA 293375
QY 580 GlyProLysSer-----IleGluAspSerSerGluMetLysLeuValArgTyr 595
Db 293376 AATCCAATACAAATGGAGCATTGTGTCARAAAATTCAGATATCAA----- 293423
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
Db 293424 AACATACTAATGCGCGCTTTAAGCGGTATTGAATTGCAAGCTCAATACCAA---ACAGAA 293480
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 293481 CGTTTAAGCTA-----TTTACTAACTATGGCAGC 293510
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 293511 ACCAAAGGTAAGATAAGATAGTGGC-----GAAGCT 293543
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
Db 293544 TTATCAACACTTGGCGCAAGCAAAATCGCGTAGGGGTAAATTTATGCTTTAGTAAAGAC 293603
QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 293604 AAATTCAGTGGGAGCGACAGTAACCCATTACGCT-----GTCACAGCGCGATGGCT 293657
QY 693 ArgTyrGluThrArgThr---ProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 293658 AAAGATCATAGTGTACTACCCCAAGTTATATCTACCGGATCTTCGGCTACCTATGCCA 293717
QY 712 ArgAsnThrArgTyrGlyGluTrp 719
Db 293718 -----CCATTAAAGGCGAATGG 293735
RESULT 10
US-09-252-991A-1375
; Sequence 1375, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18

QY		231	ArgArgPheAlaasnGlyGlnHisargAlaValLeuGlyTyrArgLysArgPheTyrArg	250
Db		809	---CGGACGGCAACAACGACAGCATATCGACGCTGCTCAGGCCATTGGCAC	865
QY		251	ArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAsp	270
Db		866	CTCACCCAGACGCACCAG-----TTGGCGGCCAACACTTCCTCACTACTAC---	907
QY		271	AspCysHisAlaAspIleIleTrpGlnLysSerLeulleAsnLysArgTyrLeuGlnLeu	290
Db		908	GAGCGCTACGCCGATATG-----	925
QY		291	TyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGly	310
Db		926	---CCGGCGGCTGACCCAGCGCAGTAGCAC--GACGATCCCTTCCAGTCGTA---	976
QY		311	PheHisAspAspAspAlaHisAlaHisAlaHisAsnGlnLysProTrpIleAspLeu	330
Db		977	-----CGCAGCTGGGACAATTC	994
QY		331	ArgAsnLysArg-----	334
Db		995	CGTGTCGGCGCAAGGACTTCTCGTGAAGTACACCGCCAGCTCGACGACCTCACCCAG	1054
QY		335	TyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis	354
Db		1055	TTGAGGTGCTCACTACTACAGCAGAGTTCCGGCGCAGCAGCATCGCGCGCCAAC	1114
QY		355	LeuAsnArg-----AsnAspTyrHisHisAspGlulylsAlaGlyAspAla	369
Db		1115	CTCAGGACCATCACTCGTACCCGCGCAGTACCATGTGTTCGCGGTGGAGCCGGGTC	1174
QY		370	ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro	389
Db		1175	TCGCGGATCTTCTTCGCGCGCCGACC-----CAGGAG	1210
QY		390	IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu	409
Db		1211	ATCGGCTACCGTACTCTGAAGA-----	1240
QY		410	SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn-----	424
Db		1241	--CGATGACAGAGCGCGCAGCAGCAGCTGGCTGTGTCAACATGTGCCGCGTGGT	1297
QY		425	-----LysValGlnHisTyrSerPhe	431
Db		1298	CCCGCTCGAGCGCCATACTACAGGACCGTACCGCGGTACCGGCGCAGCGCCCTTC	1357
QY		432	PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgValGlu	451
Db		1358	TACATCGACACAGATCGATGTGCGCAATGGACCATCACCCCGGCATCCCGTTCGAG	1417
QY		452	LysGlnLysAlaSerIleArgTyrAspLysAlaLeulle-----	464
Db		1418	AAGATCAGCAGCAGTGGCGC--GACCGCCGCTGCTCGCGCTGAACGCGACCGCGTG	1474
QY		465	-----AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg	481
Db		1475	CAGGAGAAGAGCCGACCAAGGACTACAAAGAACCGCTGCGCGCTG-----	1522
QY		482	GlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu	501
Db		1523	-----AGCCTGATGTATCACTGAGCGCAGCATGG-----	1552
QY		502	SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGly	521
Db		1553	AAGCTGTGCGCAACTACTCGAATCTGCGCAGCCTGCAG-----TACTTCCAGCTCGC	1609
QY		522	LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer	541
Db		1610	CGGCGCGCGCAACGATACCGCGCCCGCG-----CTGAGCGCGGAGGAGGCC	1660
QY		542	AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeu	561

1039 AATTGGCGTACACCGCAAAAGATTGGCTTTGTCTGCAAAAGTGGATGCCAACGCTAT 1098
 321 AlaHisasnGlyLysProThrPheLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
 1099 GTGCT---GGAATAAAAGCGTATTTCGGCGGATGACAAAGATACACGGCTACGCGAGCAA 1154
 341 TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
 1155 TGTAAAGAGGCGCAACCA-----TACCGGATCGCCACTCGGGG 1193
 361 HisHisaspGluLysAlaGlyAspAlaValIleuAsnPhePheAsnAsnGlnThrGlnAsn 380
 1194 CAT----- 1196
 381 AlaArgIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValGlu 400
 1197 ---GAACCTTCAACTTCACAGCGCGCTTTCGGCAACAAACCTTTGAAATACGGCATCAA 1253
 400 nTyrLeuGlyGlnLys-----SerSe 407
 1254 CTACCGGCATCAGGAATCAACCGCAGCGCTTTTGAATTCACATTTAAATTTGAAGA 1313
 407 rAlaLeuSerAlaThrSerGluAlaValLysGluProMetLeuLeuAspAsnLysValGlu 427
 1314 TAAATAAGATGCAACTGAGGAAGATAAATGGAAGAACCGTGCAAAATGAAAAAATTGCCAA 1373
 427 nHisTyrSerPhePheGlyValGluGlnAlaAsnTrp----- 439
 1374 AGCTTACGCTGTACCAACCGCCGACCAACCGATACCGCGCTATATCGAAGCCATTCA 1433
 440 -----AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIl 457
 1434 CGAGATTGACGGCTTTACCTGTACCGCGCGCTCGTTTACGACCGCTTCAAGGTGAAAAAC 1493
 457 eArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys---GlnProLeuProAs 476
 1494 CCAC-----GACGG-CAAAACCGTTTCAAGCAGCAGCGCTCAACCGCA 1534
 476 pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT 496
 1535 GTTT-----CGCGCTGATTGGCAGCGCGCGCGCAACACACTGGAGCTTCA 1576
 496 hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG 516
 1577 GCGGAGGCAACAATAC-----GCCAGCGGAGCGCGCGCTG-----TATGACG 1621
 516 luLeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValG 532
 1622 CGCTGCAAAACCGCGCAACCGCGCATCATCTCGATTGCGGACCGGAC 1671
 532 lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG 552
 1672 -----AAACCGCAACCGCGCGCATACCGGAATACCGAATCGGCTTCACTACAACG 1717
 552 lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg----- 563
 1718 ACGGACGCTTTCGCGCAACCGCGCTTTCGCGCAGACCATCAAGACGCGCTTGCCA 1777
 564 -----AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
 1778 ATCCGCAAAACCGCGCAGCTGTGCGCGTCCGCGAAGCGCTCAACCGCGCTTACATCA 1837
 578 ----- 578
 1838 AAAACACCGTTACGAATTGGCGGCTCTTACCGCAGCGCGGCTGACCGCAAGTCG 1897
 579 -----ArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrA 596
 1898 GGTAAAGCGCGCAACCGCGCTTTTACGATACCATCCCTCAAAAAAACTGTTGAGCGCGA 1957
 596 snGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyGluIleTyr----- 610
 1958 ACCCGAG-----TTTGGCGCAACACCGCGCGCATTTGGACGGCTCCCTTG 2005

611 -----PheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG 628
 2006 CCTACCGCTTCAAA---AACCGAATCTGGAATCGCTGGCGGAGCGCTATGTT--- 2058
 628 lyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProp 648
 2059 -----CAAAAGCTACGGGTTGATATTGG 2093
 648 heIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL 668
 2094 CGGAGGCGCAAAAGACCGGACCGCAAAATTGGAACCGTTGTACGCCACGAGTTTCGGTG 2143
 668 euLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG 688
 2144 TGAAC-----GATGCTCTTCGCCA 2161
 688 lnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyA 708
 2162 ACTGGAACCGCTGGCGCAAGACACGCTCAAT-----GTTAATCTTCGG 2206
 708 laAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnL 728
 2207 TTAAC-----AACG 2215
 728 euLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetG 747
 2216 TGTTCGACAAGTTCTACTATCGCACAGCAACGCTGGACCAATACCTGCGCGCGGTGG 2275
 747 lyArgSerPheThrGlyGlyValAsnValLysPhe 758
 2276 GACGTGATGATCGCTGGCGTGAACCTCAAGTTTC 2310

RESULT 12

PCT-US96-05320A-113
 ; Sequence 113, Application PC/TUS9605320A
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences
 ; APPLICANT: 9410 Key West Avenue
 ; APPLICANT: Rockville, MD 20850
 ; APPLICANT: United States of America
 ; APPLICANT: Johns Hopkins University
 ; APPLICANT: 720 Rutland Avenue
 ; APPLICANT: Baltimore, MD 21205
 ; APPLICANT: United States of America
 ; APPLICANT: Mark D. Adams
 ; APPLICANT: Owen White
 ; APPLICANT: Hamilton O. Smith
 ; APPLICANT: J. Craig Venter
 ; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue, Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20003-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/05320A
 ; FILING DATE: April 22, 1996
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/476,102
 ; FILING DATE: June 7, 1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/487,429
 ; FILING DATE: June 7, 1995

Qy	226	LeuProAspSerProArgArgPheAlaasnGlyGlnHisArgAlaValLeu	----- 242
Db	646	CCCGATAATAGCAAAATCTCTTTTCAAAAAATAATCAAAAAACCGGATTAATAAAATTA	705
Qy	242	-----	242
Db	706	AATTGGCAAAATTCACCGACATTTTACTACATTATCGAGTGTATTGTTATGTTATTCATATAA	765
Qy	243	GlyTirArgLysAlaGpHeTyArgArgThrTyrSerAspArgArgAspGlnTy-GlyLeu	262
Db	766	GGCTGG-----GAGCCYTTCCGAGCAAAAAGATATCCTACCTATAA	807
Qy	263	ProLalHisSerHis-----GluTyAspAspCysHisAlaAspIleIleTirpGlnLys	280
Db	808	CCAAAGTTTGAGTGATATAATCGTGTATGGT-----ACTGATATAGCGTGGAAACGT	858
Qy	281	SerLeuIleasnLysArgTyLeuGlnLeuTyProHisLeuLeuThrGluGluAspVal	300
Db	859	AAACTGGTTATCGA-----GATCAAAAAGATGAA	888
Qy	301	AspTyAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
Db	889	AAATTACATTA-----AAATCAACATAT	912
Qy	321	AlaHisasnGlyLysProTirIleAspLeuArgasnLysArgTyGluLeuArgAlaGlu	340
Db	913	CTACCTGAAAAATAACCCATGATTAACTTATCTACTCAG-----	951
Qy	341	TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr	360
Db	952	-----TTCAGCTACAGTAAAAACC-----ACACAAATATGATATG	984
Qy	361	HisHisAspGluLysAlaGlyAspAlaValIcIluAsnPhePheAsn-----	375
Db	985	CGTCCCAAGAGCATCTTCAGGATTTAGTGGTAGCTAGGTATCAAGAGCTGGATTA	1044
Qy	376	-----AsnGlnThrGlnAsnAlaArgIleGluLeuArg---	386
Db	1045	TATTTCAGATCTTACTTTTGTATATAAATAACACAGTAGTACTTTTATATATCAAAACTACTGTT	1104
Qy	387	HisGlnProIleGlyArgLeuLysGlySerTirpGlyValGlnTyLeuGlyGlnLysSer	406
Db	1105	CATGAACATA-----TTTGGTTTGCATGTGGTTAAAAAACAACAAGA	1146
Qy	407	SerAlaLeuSerAlaThrSerGluAlaValLys-----	417
Db	1147	AAATCTTTATGTATGATAAAGTAAGTGGGAAAAACGGATTAATAATTATGGCTATTTC	1206
Qy	418	GlnProMetLeuLeuAspAsnLysValGlnHis---TyrSerPhePheGlyValGluGln	436
Db	1207	CAACCCCTATTATATGCCATCTGGACGTCATATATACACAGGCGTTTTATTATACAAGATCAA	1266
Qy	437	AlaasnTrpAspAsnPheThrLeuGluGlyGlyValArg-----Val	450
Db	1267	ATAAAATGGAAGAATATAAATTTTATAGCACAGGAGTGGTTATGACCATATCAATAATA	1326
Qy	451	GluLysGlnLysAlaSerIleArgTyAspLysAlaLeuIleAspArgGluAsnTyTyr	470
Db	1327	GGGCAGAAAAATTAGCACTAAAATNTAATGAT-----ATCTCTCAGGACATGATAT	1380
Qy	471	LysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer	490
Db	1381	AGCCAG-----AAAAATTAAATGTTGGTCTTATTATTATAGGT	1419
Qy	491	GlyAsnTrpTyrrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu---	509
Db	1420	TTAAT-----TATGATGTAAACCATATTATTAAGTTTATTACGAATTTTAGTAAAACT	1473
Qy	510	---ArgLeuProSerThrGlnGluLeuTyrrAla-----HisGlyLysHisValAla	525
Db	1474	TGCGGAGCCTGTTATTGATGANCAGTATGAGACGCAATTTAAGCAATCTTCTGTGCCT	1533
Qy	526	ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu	545

Db 1534 GCACCTCTTAAATTTAGAAAAGAAATGATTAATCAACGACAGTGGGTGAATT--- 1590
Qy 546 LeuAlaLeuGlyTyr-----GluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyr 562
Db 1591 ATTACTCTCAATCATCTATTTCAGGAAATATGATCTTTTCAATTAGAACTATTATTT 1650
Qy 563 ArgAsnArgPheGlyAsnTyrIleTyr-----Ala 572
Db 1651 TACAATCGCGCAAGAATGAATATTTAAACACGAGGAGTAAATGTGTAGGCAATGCT 1710
Qy 573 GlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeu 592
Db 1711 GCAGATACCAATAATAGGTTTGTCTTAATCATTCGAAT- 1752
Qy 593 ValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLys 612
Db 1753 ---TATCGTAATTTGCCAGGTTATGTTATTCAAGGTGCAGAGTTAGAAGCTTATTATCAA 1809
Qy 613 ProThrProArgTyrArgIleGly---ValSerGlyAspTyrValArgGlyArgLeuLys 631
Db 1810 TCTACT-----TATTATTGGTGAGATAACATATTCTTATGTAAAGGAAACGGTAC 1863
Qy 632 AsnLeuProSerLeuPro---GlyArgGluAspAlaTyrGlyAsnArgProPheIleAla 650
Db 1864 ACCTCACCAGAAACCCATGGGGTAAACCTCAACATGG-----ATTGCT 1908
Qy 651 GlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAla 670
Db 1909 -----GAAATPCCCTCCCTAGAAAACCACTACTGCTTTAGGTTTCAATGTTCCAAA 1959
Qy 671 SerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLys 690
Db 1960 TAT-----TATCTCAGAGTAGGTGGCGGTGAG 1989
Qy 691 LeuAlaArgTyrGluThrArgThrPro-----GlyHisMet 703
Db 1990 TTTGTAAAGACAGCAAGATCGATCGCCATTATCTGGTGATCCTAAAGCATCATCTTGTC 2049
Qy 704 LeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyr----- 719
Db 2050 TTGCCAGCTTCAAGAGATACAGCCTACATAACCTATTCTCTTGGAGCCCTGCAAAA 2109
Qy 720 -----AsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAla 735
Db 2110 ATTAAAGGTATGATGTAGATTACAGTGTGATTAATTATTCATCAGCATAC----- 2163
Qy 736 HisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrGly---GlyVal 754
Db 2164 -----AACCTTATCTAGCGAATTGGCTTCTGGAACAGGTAGG 2202
Qy 755 AsnValLysPhe 758
Db 2203 AATATCAATTT 2214

RESULT 13
US-09-252-991A-12930/c
; Sequence 12930, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12930
; LENGTH: 618

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12930
Alignment Scores:
Pred. No.: 2.82e-13 Length: 618
Score: 216.00 Matches: 64
Percent Similarity: 45.13% Conservative: 38
Best Local Similarity: 28.32% Mismatches: 90
Query Match: 5.35% Indels: 34
DB: 4 Gaps: 6

US-09-936-377-2 (1-758) x US-09-252-991A-12930 (1-618)

Qy 506 SerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValala 525
Db 616 AGCTACACCGGCGCGCACCGAGCTTCTACGAGCTGTACGCCACGCTCCGACGCCGCC 557
Qy 526 ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnIleGlu 545
Db 556 ACCGGCACCTACGAGGTAGCGATGCCGACGACGAAGGAAAGGGGGTCTCCACCGAC 497
Qy 546 LeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArg 565
Db 496 CTCGCCCTGCGCTTCGACAAACGGCGTGCACAAAGGCGAGCTGGGGTGTCTACAGCGC 437
Qy 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLys----- 582
Db 436 TTCTCCAACTACATC-----GGCCTGCTGCCGACGCGTCCCATCGCAACGAGGAAGGC 383
Qy 583 -----SerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 382 GAAGTGGTTCGCCCGCACCGATGACGAGCGCTGCGGATACCTCTACAGCGCGTTTCGC 323
Qy 600 AlaAspPheTyrGlyAlaGluGlyGlu-----IleTyrPheLysProThrPro 615
Db 322 CGGACTTCTACGCGCTCGAGGCGCAGACCGCATCCACTGTGGAAGCCCGTACGGC 263
Qy 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 262 AACTTCGACCTGGAACCTCTCCGGGACTACACCGGACCAAGAACAGGAC----- 212
Qy 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 211 -----ACCGCGCAACCG 200
Qy 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArg 675
Db 199 CTGCCACGATCCCGCGCTGCCCTGGAACACCGCGCTGATCTGGGAGTTG---CAGCAG 143
Qy 676 IleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGlu 695
Db 142 TGGCAGCGCGGGTGCAGCTCGAACACGCGCTCGCAGCACCGCGTCCGCGGAGGAAGA 83
Qy 696 ThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArg 715
Db 82 CTCTCCACCGACGCGCTAC-----ACCACCTCGCGCGCAGCTCGGCTCAACTTCGAC 29
Qy 716 TyrGlyGluTyrAsnTrp 721
Db 28 CTGCGCGAGAGCGCGCTGG 11

RESULT 14
US-09-252-991A-6426
; Sequence 6426, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

239	ArgAlaValLeuGlyTrp	---244
766	GAGCTCGCGTGGCTGGACCCCGACGAGCACACCTGATCGAACTACCGCGCGCAAG	825
245	-----ArgLysArgPheTyrArgArgThrTyrSerAsp	259
826	GGCGACGGCGAGCAAGCTACGCGCGCGCGGATGATGACGGCTTCGAGTTCGAAGCGCGAG	885
259	GlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleLeuTrp	278
886	AGCCTGGCGCTGGCTTCGTCAATCGAAGCTCAGCGAT	924
279	GlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu	298
925	-----GTGCTCGAGAAGGTCGAGGCACAGGTTTAC	954
299	AspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHis	318
955	-----TACAATTACCCCGAC	965
319	AlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArg	338
970	-----CACATCAGG	978
339	AlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn	358
979	GACAAGTTCGCGCTGGCACTCCCGGA	1017
359	AspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThr	378
1018	GATGCCATCCCGATGGCGTCCAGGT	1053
379	GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer-TrpGln	398
1054	CACCTCGCGGTGGCTGGCGCCACTTGGCGTGGAGCACTTCAAGTGGTCCACCG	1113
398	YValGlnTyrLeu-----GlyGlnLysSerSerAlaLeuSerAl	411
1114	GGTGGACGCATCGCAACGAGCACCGCGCGCGCTCCAAAGTACGACATGATGACCGGA	1173
411	aThr--SerGluAlaValLysGlnProMetLeuLeuAsnLysValGlnHisTyrSe	430
1174	NTANTACACGCGCGACCACTCCCTCGAGCAGGACCGGTGTCCACAACACTACGG	1233
430	rPhePheGlyValGluGlnAlaAsnTrp-----AspAsnPheThrLeuGluGlyGln	447
1234	GGCGCTTCGGC-----GAACTGACCTGGTTCGCGCGCGACCGCGCTGATCGCGCG	1287
447	YValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGln	467
1288	GTTCGCGCTCGAC-----CGCGCTCGGTCAAG-	1315
467	uAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPh	487
1316	-----GACTACCGCGACGACCTGGAAGTCCGG	1341
487	eAlaLeuSerGlyAsnTrpTyrPheThrPro-----	497
1342	GCACATG-----GGCACGCCATGGCCAACCGCGCAACCGCGCGCGATACCCCT	1398
498	-----GlnHisLysLeuSerLeuThrAlaSer-----	506
1399	GCCACGGCTTCGTGCGCTACGAGCAGCACTCGCGACTCGCGCGACCCCTATACGC	1458
507	-----HisGlnArgLeuProSerThrGlnGluLeuTyrAla-----HisGln	521
1459	CGGCTCGGCATCGCCAGCGCTTCCCGACTACTGGGAGCTGTCTCGCCCAAGCGCGG	1518
521	YLysHisValAlaThrAsnThrPheGlnValGlyAsnLysHisLeuAsnLysGluArgSe	541
1519	GCGGACGGGTGGTCAACGCTTCGAC-----AAGATCAAGCCCGAAGAC	1566
541	rAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLe	561

299	AspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHis	318
955	-----TACAAATTACCCCGAC-----	969
319	AlaHisAlaHisAsnGlyLysProTrrPleAspLeuArgAsnLysArgTyrGluLeuArg	338
970	-----CACATCAGG	978
339	AlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnAspAsn	358
979	GACAAAGTTCGCGCTGGGCATCCCGGA-----CCCATCCGGCAT	1017
359	AspTyrHisHisAspGlnLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThr	378
1018	GATGCCATCGCGATGGCGTCCCGGT-----GGACAGAGG	1053
379	GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer-TrpGln	398
1054	CACCCTCGCGGTGGCTGGCGCCACTTGGCGCTGGGACACTCAAGTGTGCACCGG	1113
398	YValGlnTyrLeu-----GlyGlnLysSerSerAlaLeuSerAl	411
1114	GGTGGACGCATGGCAACGAGCACCGCGCGCGCTCCCAAGTACGACATGATGACCGA	1173
411	aThr--SerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSe	430
1174	NTANTACACGACGCCGACAGCACTTCCCTCGAGCAAGACCGGTGTCCACAACTACGG	1233
430	rPhePheGlyValGluGlnAlaAsnTrp-----AspAsnPheThrLeuGluGlyGln	447
1234	GGCCTTCGGC-----GAATGACCTGGTTCGCCGCGGACCGCGCTGATCGCGCG	1287
447	YValArgValGlnLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGln	467
1288	GTTGCGCTGGAC-----CGCGCTCGGTCAAG-----	1315
467	uAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPh	487
1316	-----GACTACGGGACAGACCTCGAGTCCGGTTCAGTCCGG	1341
487	eAlaLeuSerGlyAsnTrpTyrPheThrPro-----	497
1342	GCACATG---GGCCACCCCATGGCCACACCGCCGACACCGCGCGCATACCT	1398
498	-----GlnHisLysLeuSerLeuThrAlaSer-----	506
1399	GCCACGGGCTTCGTGGCTACGAGCAGACCTCGCGCATCGCGGACCCCTATACGC	1458
507	-----HisGlnGluArgLeuProSerThrGlnGluTyrAla-----HisGln	521
1459	CGGCGCTCGGCATCGCCGAGCGCTCCCGCACTACTGGGAGTGTCTCGCCCAACGCGCG	1518
521	YLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSe	541
1519	GCGGACGGGTGGTCAACGCTTCGAC-----AAGATCAAGCCGCGAAGAC	1566
541	rAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLe	561

[illegible]

```

;
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 318..2456
; US-08-628-434-1
Alignment Scores:
Pred. No.: 2,58e-11 Length: 2600
Score: 207.00 Matches: 196
Percent Similarity: 34.66% Conservative: 100
Best Local Similarity: 22.95% Mismatches: 308
Query Match: 5.13% Indels: 254
DB: 3 Gaps: 41
US-09-936-377-2 (1-758) x US-08-628-434-1 (1-2600)
QY 11 LeuSerIleLeuLeuIleAsnThrProLeu---LeuAlaGlnAlaHisGluThrGluGln 29
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
339 CTGACGCTGCTCTCGCTCACACTTGCCTCCCGCGCTTTGCCACGCGCGCAGAAAATAATGCC 398
QY 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
399 AATGTCGATGGATACCGTTACCGTAAAGCGCACCCGCAA----- 440
QY 50 LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLys 69
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
441 -----GGCAGCAAAATCCGTACCAACATCGTT-----ACGCTTCAACAAAA 482
QY 70 -----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
483 GACGAAGCACCAGCAACCGATATGCGGAATCTTTAAAGAGAGAGCCCTCCATC----- 536
QY 86 SerGlnTyrGlyGlyAlaSerAlaProValIle-----ArgGlyGln 100
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
537 ---GATTTCGGCGGGCGCAACGGCAGCTCCCAATTCCTGACGCTCGCGGTATGGGTGAG 593
QY 101 ThrGlyArgArgIleLysVal-----LeuAsnHisHis 111
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
594 AACTCTGTCGACATCAAGGTGGACACACGCTATTCCGACAGCAAAATCCTTTACCACCAA 653
QY 112 GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla 131
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
654 GGCAGA-----TTTATTGTGATCCCGCT 677
QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
678 TTGGTTAAAGTCGTTTCGTACAGAAAGC-----GCGGTTCC 716
QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu----- 169
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
717 GCCTCTGCGGTATCGCGCGACCAACGCGCGCATCATCGCAAAACCGTCGATGCCCAA 776
QY 170 -----AsnGlyValSerGlyLeuLeuGlyLeuArgLeuSerSer----- 182
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
777 GACCTGCTCAAGGCTTGGATAAAACTGGGCGCGTGGCGCTCAACAGCGGCTTTGCCGCGC 836
QY 183 -----GlyAsnLeuGluLysLeu 188
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
837 AACAAACGCGCAAGCTACGGCGCAAGCGTATTCGGAAGAGGGCACTTCGACGGTTTG 896
QY 189 ThrSerGlyGlyIleAsn-----IleGlyLeuGlyLysAsnPhe-ValLeuHi 204
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
897 TTCTCTTACAAACCGCAACGATGAAAAAGATTACGAAGCGGCAAAAGGTTTCCGCAATGAC 956
QY 204 sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLy 224
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

RESULT 15

```

US-08-628-434-1
; Sequence 1, Application US/08628434
; Patent No. 6265567
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule and Vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,434
; FILING DATE: 05-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/418,964
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-3-P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405

```

957 AACG-----GGCGCAAAACCGTACCGTACAGCGCGTGGCAAAACGCGAGC 1001
 224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
 1002 TACCTCGCAAAATCGGAACAAC--TTGGGAGCGGGGACCAACCGCATCGTGTTGAGCCA 1060
 244 pArgLysArgPheTrpArg-----ArgThrTyrSerAspArgAspGlnTyrGlyLe 262
 1061 TATGAAGACCAACACCGGGCATCCGCACTGTG-----CGTGAAGAGTTTGGCGT 1111
 262 uProAlaHisSerHisGluTyrAspAspCysHisAlaAlaPheLeuTrpGlnLysSerLe 282
 1112 CAGCGCAAAAT-----TACCGGAT 1132
 282 uLeAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
 1133 AACTATTAAACCGCAAGCCCATCTACCGGAA--ACCATTCAATCAACCAACAACCT 1189
 302 rAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
 1190 GCGGTACACCGCAAGAGTTTGGCTTTGTGCAAAACCTGGATGCCAAGCTATGTGTT 1249
 322 sAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLys 342
 1250 ----GGAAAAGAAACGCTATTCCGCGCATGACAAGATTAACGGCTACGAGCAATGTAA 1305
 342 sGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHi 362
 1306 AAGGCCCAACCA----- 1318
 362 sAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThrGlnAsnAlaAr 382
 1319 -----TACCGATCGCCACTCGAGTATGA 1344
 382 gIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValGlnTyrL 402
 1345 ACTTCAACTTCAGACCGCCCTTGGCGAACAACCGTGAATAATGCAAAAGCT 1524
 402 euGlyGln-----LysSerSerAlaLeu----- 409
 1405 GCATCAGGAATCAAAACCGCAAGCGTTTGTGAATCGGAAATTTGAAATAAAGATAAG 1464
 410 --SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisT 429
 1465 AAAAGCACTAATGAAGAGAAAAAGAACGCGTGAATAATGCAAAAGCT 1524
 429 yrSerPheGlyValGluGlnAlaAsnTrp----- 439
 1525 ACGCCTGACCAACCGCAACAAACCGATACCGCGCGTATATCGAAGCCATTACGAGA 1584
 440 --AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgT 459
 1585 TTGACGCGCTTACCTCGACCGCGCGCTGCTGTACGACCGCTCAAGGTGAACCCAC- 1643
 459 yrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys---GlnProLeuProAspLeuG 478
 1644 -----GACGG-CAAAACCGTTTCAAGCAGCAGCGCTCAACCCGAGTTT- 1684
 478 lyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheThrPro 497
 1685 -----CGCGTGATTTGGCAGCGCGGACACTGGAGCTTCAGCGG 1727
 498 GlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeu 517
 1728 AGCCACAACATAC-----GCCGCGCGCAGCGCGCGCTG-----TATGACGCTCTG 1772
 518 TyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsn 533
 1773 CAACCCAGCGAGCGCGCATCTCTCGATTGCCGACGGCAG----- 1817
 534 LysHisLeuAsnLysGluArgSerAsnAlaGluLeuAlaLeuGlyTyrGluGlyAsp 553
 1818 -----AAAGCCGAACGCGCGCAATACCGAAATCGCTTCAACTACAACGACGCGC 1868

554 ArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln 573
 1869 ACCTTTTCCGCAAC-----GGCAGCTAC--TTCCGCGCAG 1901
 574 ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuVal 593
 1902 ACCATAAGACGCGCTTCGCAATCCGCAAAACCGCACAGCTCCGTCGCCGTCGCGAA 1961
 594 ArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyIleTyrPheLys 612
 1962 CCGCTCAACCGCGCTACATCAAAAACACGCTTACGATTTGGCGCGCTCTACCGCACC 2021
 613 ProThrProArgTyrArgIleGlyValSer----- 622
 2022 GCGCGCTGACCCGCAAGTCGCGCTAAGCCACAGCAACCGCGCTTTTACGATACGAC 2081
 623 -----GlyAspTyrValArgGlyArgLeuLys 631
 2082 AAAGCAAGCTGTGAGCGGAACCTGAATTTGGCGCAAGT-CGGCGGCACTTGGAC 2140
 632 AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGln 651
 2141 GGCTCCCTTGCCTACCGCTTCAAAAACCGCAATCTGGAATCGG-----CTGGCGCGG 2194
 652 AspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHis-LeuLysAlaSe 671
 2195 TCGTATGTTCAAAAGCCGT-----GGTTCGATTTGGCGGAGG 2236
 671 rLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLe 691
 2237 TCAAAAAGACCGC--GACGGCAATTTGGA-----AACGTTGTACGCCAA----- 2279
 691 uAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrAr 711
 2280 -----GGTTTCGCTGTAACAGATGCTTCGCCCACTGGAA 2314
 711 g-----ArgAsnThrArgTyrGlyGluTyrAsnTrpTyrValLysAlaAspAsnLe 728
 2315 ACCCTGGCAAGACAG-----CTCATGTTAATCTTTCGCTTAACAACGT 2362
 728 uLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGl 747
 2363 GTTCACCAAGTTCTACTATCGCACAGCAACGCTGGACCAATACCTCGCGCGCTGGG 2422
 747 YArgSerPheThrGlyValAsnValLysPhe 758
 2423 ACGTATGTACCGCTGGCGGTGAACATCAAGTTTC 2456

RESULT 16

US-08-425-843-1
 ; Sequence 1, Application US/08425843
 ; Patent No. 6020154
 ; GENERAL INFORMATION:
 ; APPLICANT: Hansen, Eric J.
 ; APPLICANT: Cope, Leslie D.
 ; APPLICANT: Jarosik, Gregory P.
 ; APPLICANT: Hanson, Mark S.
 ; TITLE OF INVENTION: H. Influenzae Hxub and Hxuc Genes, Proteins
 ; TITLE OF INVENTION: and Methods of Use
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210-4433
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

679	Db	CTTATTAGTGGTTTCTATAATAATTAATCGCGATAAATTTACGCACTGGTAAAGCAATAAGCTG	730
204	Qy	HisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg	221
739	Db	AAATAATACCGCC---TATAACAGTTTGGGGCTTAGCAAAATTCGGTTGCGAAATTAAT	795
222	Qy	AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg	239
796	Db	GATGCTAACCGCGTGAATATATCCACCGCGAAACTCGTTTAAACAAACAGCACCAA	853
240	Qy	AlaValLeuGlyTyrArgLysArgPhe	248
854	Db	GCAATAATAGGTGCGAAACGAACCTTACCAATGAACAAATTACAGATCAATACAGAGT	913
249	Qy	TyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSer	266
914	Db	TCCACAACCAACAAACGGTCTCCACGGAAGCCAAACCAT---CACAAAG	961
267	Qy	HisGluTyrAspAspCys---HisAlaAspIleIle	279
962	Db	AAGAGTTTACTCTGGCGTGAACACACGTTTGGTAGTGTCAGTTATTTAACTGATCAAC	1021
280	Qy	LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp	299
1022	Db	AAATTCCTGATCAAAAGCACGCTATTTTAACTAATTTAT---TTAAACGCCA---	1065
300	Qy	ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla	319
1066	Db	-----GATAATCTCTATCTA-----AATACGCATATC	1092
320	Qy	HisAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAla	339
1093	Db	GCACGTATAACAATAAACTATTGGAAGAACAGCGTAAAGTCAGTGGTGTGAAGAT	1152
340	Qy	GluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn	358
1153	Db	CGACTAAATTGACTACCCGAGGT-----ATAAATTTACGTAATTTCTTCC	1197
359	Qy	AspTyrHisAspGluLysAlaGly	367
1198	Db	GAATATCTCACATTTCCTTGGTTTATGGGGTGATTTATGCGAGATAAAATCCGTACC	1257
368	Qy	AspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHis	387
1258	Db	GAA-----CGAGGCACAAACAATAAAGATGCGAGTTTCGAGCG	1296
388	Qy	GlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSer	407
1297	Db	GACCCCTAATCGCAATTCAAACACTACAGCGGTT---TATTTA-----	1338
408	Qy	AlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGln	427
1339	Db	-----ATCGCC	1344
428	Qy	HisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGly	447
1345	Db	CATATTCGCTAATTTGGG-----GAAAAATTTGCTACTTTCCGCAAGT	1386
448	Qy	ValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGlu	467
1387	Db	GTACGTTATGACCACCTACGATACCTCA-----AGTAAAACTGTAAAAACAAGAT	1437
468	Qy	AsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPhe	487
1438	Db	AAATCATTTA-----TCTCCT	1452
488	Qy	AlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis	507
1453	Db	GCCACAAAAATTAACTTGGTAGTGACC-----AATTGGTTAGATTTTACTGCCAATAT	1506
508	Qy	GlnGlu-----ArgLeuProSerThrGlnLeuLeuTyrAlaHisGlyLysHisValAla	525
1507	Db	AATGAAGCTTTCCGAGCACCATCTATGCAAGACGGTTTGTGATGGTTCGACATTGGG	1566

QY 526 Thr-----AsnThrPheGluValGlyAsnLysHisLeuAsn 537
 DB 1567 ACAAGTATTCTAGGGCGAAATGAATCTATAAATT---GTAGCAAAATCCAAATTTGGCG 1623
 QY 538 LysGluArgSerAsnAsnIleGluLeu-----AlaLeuGlyTyr 550
 DB 1624 CTGGAACAGCGGAAATAAAGAAATACCGCAAAATCTACATTTTGATAGTCTGTTTAA 1683
 QY 551 GluGlyAspArgTrpClnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIle 570
 DB 1684 CAAGCGATAAATCAAAATGGAAGCGACTTATTTCCGTAATGATGTGAAGATTTTATT 1743
 QY 571 TyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGluMet 590
 DB 1744 AACTTAAATATTAAATGATGACAAAGCAAAATCAAAATGCAAGTGAAGTGA 1797
 QY 591 LysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluLeuTyr 610
 DB 1798 -----GGTGCAGGTGCA-----GGTCAAAATCCAAATCGAGCA 1830
 QY 611 PheLysProThr---ProArgTyrArg-----IleGly 620
 DB 1831 TTGTTGCCACAAAATCCAGTATCAAAACATAACTAATGCCCGTTTAAGCGGTATTGAA 1890
 QY 621 ValSerGlyAspTyrValArgGlyArgLeu-----LysAsnLeuProSerLeuPro 637
 DB 1891 TTGCAAGCTCAATACCAAAACAGAACCTTTTAAACACTGTTTACTAATATATGCGACCAAAA 1950
 QY 638 GlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaPro 657
 DB 1951 GGTAAAGATAAGATAGTGC-----GAAGCTTTATCA 1983
 QY 658 ArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp----- 674
 DB 1984 AACATTGCCGCAAGCAAAATCGCGTAGGGGTAAATTTATGCTTTAGTAAAGACAAATTC 2043
 QY 675 ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyr 694
 DB 2044 ACGGTGGGAGCAGACGTAACCCATTACGCT-----GCTCAAGCGCGAGTGCTTAAAGAT 2097
 QY 695 GluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsn 713
 DB 2098 CATAGTGTACCTACCAAGTTATATATCTAGCCGATCTTCGTCTACTATGCA----- 2151
 QY 714 ThrArgTyrGlyGluTrp 719
 DB 2152 CCATTAAAGCGAATGG 2169

RESULT 17

US-09-252-991A-14350
 ; Sequence 14350, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 14350
 ; LENGTH: 1854
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-14350

Alignment Scores:

Pred. No.: 1.29e-10 Length: 1854

Score: 198.00 Matches: 157
 Percent Similarity: 34.26% Conservative: 90
 Best Local Similarity: 21.78% Mismatches: 254
 Query Match: 4.91% Indels: 220
 DB: 4 Gaps: 39
 US-09-936-377-2 (1-758) x US-09-252-991A-14350 (1-1854)
 QY 7 LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
 DB 70 GAACCGGTGAGCTTGGCCGACCGAGTGTGGCCGACCGCTACCGCCGAGCC----- 123
 QY 27 ThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAla 46
 DB 124 -----GCTTCCAGAGCCTGGCGCGGTGAGCGCTCATCGACCGCGAG----- 165
 QY 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu 66
 DB 166 -----GATATCGAG 174
 QY 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
 DB 175 GCGAGCCAGCGCGCGAGCTGTGGCCGAGGTGTGGCCGAGGTACCGCGGTGTGCTGGCG 234
 QY 87 GlnTyrGlyGly---GlyAlaSerAlaProVal---IleArgGlyGlnThrGlyArgArg 104
 DB 235 AACAAACGCGGTTCGCGCAAGAACACACCGCTGTCTGCGCGGACCGAGTCCGACCAT 294
 QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAsp---MetAlaAspPheSerProAsp 123
 DB 295 GTGCTGGTGTGATCGAGCGGCATCAAGTCTGGCTCGGCCAGCGCTCACCCTCACCCTGTC 354
 QY 124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
 DB 355 CAGGACTTTCGCGGTGGAG-----CTGATCGAGCGCATCGAGGTGTTCGCGCGCGCGCT 408
 QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal-----Ala 159
 DB 409 TCCAGCTTGTAGCTCGAGCGCATCGCGCGGTGTATCCAGATCTTCACCCGCGCGCGC 468
 QY 160 AspGlyLysIleProGluLysMetProGluAsnGly-----ValSer 173
 DB 469 GACGGCCAGGGCGCGCAAGCGCTTCTCTCCGCGGTACGGCACCCATCAGACCTGGAG 528
 QY 174 GlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIle 193
 DB 529 GCGAGCGCGGGTCTCGCGCGCGCGCGCGCAAC-----GGCTGGTAC 570
 QY 194 AsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys----- 211
 DB 571 AGCCTCGGGTGGAGCAGC-----TTGATATCGCGCGGGATCAATACCNAGCGCGCC 621
 QY 212 ---SerGlyAspTyrAlaValProArg-----TyrArgAsnLeuLysArgLeuProAsp 228
 DB 622 GGTACTCGGGGTATGACCGCAGACCGAGCGCTACCGCAACCTGTCGCGCAACCTGGCG 681
 QY 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 248
 DB 682 GCGCGCTATCGCTTCGACAAATGGCGCTG----- 708
 QY 249 TyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 268
 DB 709 -----GAATCGACGCGACCGCTGCTCAGGGCCCAAGTCCGCAAC 747
 QY 269 TyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeu 288
 DB 747 ----- 747
 QY 289 GlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSer 308
 DB 748 -----GACTATGACCGAGTTTTCGGCAAC 771
 QY 309 CysGlyPheHisAspAspAlaHisAlaHisAsnGlyLys----- 325

```
Db 772 TCCGGTTTCAATGCCAACCGGACGGGACGACGAACTGGTGGCGCGCTGCCCGCTTC 831
Qy 326 -----ProTrp-----IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
Db 832 ACTCCGTTCCGATCCCTGGCTGGTGACCTCCAGCGCGGGCGACGAGGACAGGCGCAT 891
Qy 341 TrpLysGlnPro-----PheProGlyPheGluAlaLeuArgValHisLeuAsn--- 356
Db 892 GCCTATCAGGATGCCCGTTTCTACTCGCGTTTGATACCGCTCGCGACAGCCTGTCTCG 951
Qy 357 ArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsn 376
Db 952 CAGAACGACCTGACCTGGCGGAA-----GGCATGTACTGACCTCGGCTACGACTGG 1005
Qy 377 GlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer 396
Db 1006 CAGAAGGACGAGATCAGCAGCAGCAGGACCTTACGCTGACCTCGCGCTGCAACAAAGGC 1065
Qy 397 TrpGlyValGlnTyrLeuLysAsnLysValGlnHisTyrSerPheGlyValGlu--- 435
Db 1066 TGGTTCGCCAGTACTCTGCGCAGTACGGT----- 1095
Qy 417 LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu--- 435
Db 1096 CGCCAGGATTGGCAACTGAGCTCGCGCGTGACGACCAACGACGAGTTTCGGCGTGCAAGC 1155
Qy 436 -----GlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGln 453
Db 1156 ACCGCGACGCGCGCTGG---GGTACGCGCTGAGCGACGCGCTGCCGCTTC--- 1203
Qy 454 LysAlaSerIleArgTyrAsp-----LysAlaLeuLeuAspArgGluAsnTyrTyr 470
Db 1204 -----ACCGTCAACTACGCGCAGCGCAATCAAGCGCGCGACCTTCAACGAACTCTACTAC 1257
Qy 471 LysGln---ProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
Db 1258 CCGGACTACGGCAATCCCGACTGACCGCGAGACTTCGCGCAGCCTCGGAGGTGGGCTG 1317
Qy 490 SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
Db 1318 AGCGGT----- 1323
Qy 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGly---LysHisValAlaThrAsnThr 528
Db 1324 -----ACGATGGCTGGGGGCACCTGGCGCGTGAATGCC 1356
Qy 529 PheGlu-----ValGlyAsn----- 533
Db 1357 TTCCGTACCAAGCTCGACGACCTGATCGGCAACGATCGCGCTCCGGGCGCGCCC 1416
Qy 534 -----LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
Db 1417 TGGGGGACCGCAACCAATCGACGAGCGCGCATCCGCGCGTGCAGACTGGTCTCTCGC 1476
Qy 550 TyrGluGlyAspArgTyrGlnTyrAsnLeu---AlaLeuTyr-----ArgAsnArg 565
Db 1477 AGCCAGTGGCTGGGTGGAGTGGAAACGCAACGCGACCTTCTCGACCCCGCAACCGT 1536
Qy 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly-----ArgGly 580
Db 1537 TCCGCGCGC-----GTCACGACGCGCAACGAGCTCCGCGCGCGCGCG 1578
Qy 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAla 600
Db 1579 CGCGGATGTTCAACTGGAATGACCGCGCGCTTCGAGCGTCTTCG---CTGGGCGCC 1635
Qy 601 AspPheTyrGlyAlaGluGlyGluLeuTyrPheLysProThrProArgTyrArgIleGly 620
Db 1636 AGCGTGACAC---GCCGAAGCGCGACGCTATGATGATCGCGGCAACAGAGTGGCGCTGGC 1692
Qy 621 -----ValSerGlyAspTyr-----ValArg 627
Db -----ValSerGlyAspTyr-----ValArg 627
Db 1693 GGCTACGCCACCTCGACCTCGCGCAGCGAGTACCGGTGAACGACGAAATGGCGCTCGAG 1752
Qy 628 GlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly---AsnArg 646
Db 1753 GGCGGATCGCAACACCTGTTCGGT---GCCGACTACGAAACCGGTATGGCTACACACAG 1809
Qy 647 Pro 647
Db 1810 CCT 1812
RESULT 18
US-09-252-991A-14267/c
; Sequence 14267, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14267
; LENGTH: 2943
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14267
Alignment Scores:
Pred. No.: 2,75e-10 Length: 2943
Score: 198.00 Matches: 157
Percent Similarity: 34.26% Conservative: 90
Best Local Similarity: 21.78% Mismatches: 254
Query Match: 4.91% Indels: 220
DB: 4 Gaps: 39
US-09-936-377-2 (1-758) x US-09-252-991A-14267 (1-2943)
Qy 7 LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
Db 2673 GAACCGGTGAGCTTCCCGACCGAGTGGTGGACCGCACCCTACCGCCGAGACC----- 2620
Qy 27 ThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAla 46
Db 2619 -----GCTTCCGAGAGCTGGCGCGGTGAGCGTTCATCGACCGCGAG----- 2578
Qy 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu 66
Db 2577 -----GATATCGAG 2569
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db 2568 CGCAGCGAGCGCGCGAGCTGGCGGAGCTGTTGGCGCAGTACCCGCGCTGCTGGCGG 2509
Qy 87 GlnTyrGlyGly---GlyAlaSerAlaProVal---IleArgGlyGlnThrGlyArgArg 104
Db 2508 AACAAACGCGGTTCGCGCAAGAACACACCGCTGTTCCTCGCGGCGACCGAGTCCGACCAT 2449
Qy 105 IleLysValLeuAsnHisHisGlyGluThrGlyAsp---MetAlaAspPheSerProAsp 123
Db 2448 GTCTGTGTCTGTATCGACGCGCATAGTTCGGTCCGCCAGCGCTGGCTCACCGGTTTC 2389
Qy 124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLeuArgGlyProVal 143
Db 2388 CAGGACTTGGCGGTGGAG-----CTGATCGAGCGCATCGAGTGTCCGCGCGCGCT 2335
Qy 144 ThrLeuLeuTyrSerGlyAsnValAlaGlyLeuValAspVal-----Ala 159
Db 2334 TCCAGCTGTACGCTCGGAGGCGCATCGCGGCGGTGATCCAGATCTTACCGCGCGCGC 2275
```

QY 160 AspGlyIleProGluLysMetProGluAsnGly-----ValSer 173
 Db 2274 GAGGGCCAGGCGCCAGCGGTCTTCTCGCGGTACGGACCCATCAGACCCTGGAG 2215
 QY 174 GlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyLe 193
 Db 2214 GGCAGCGCGGGGTACAGCGCGCGCGCGCAAC-----GGCTGGTAC 2173
 QY 194 AsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuThrArgLys----- 211
 Db 2172 AGCTCGGGGTAGCAGC-----TTGATACGGCGGGGATCAATCCAGCGCGCC 2122
 QY 212 ---SerGlyAspTyrAlaValProArg-----TyrArgAsnLeuLysArgLeuProAsp 228
 Db 2121 GGTACTCGGGCTATGAGCCAGACCGAGCGGTACCGCAACCTGTCGGCAACCTCGCG 2062
 QY 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 248
 Db 2061 GGGGGCTATCGCTTCGCAATGCGCTG----- 2035
 QY 249 TyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 268
 Db 2034 -----GAACTCGACGGCAGCGTCTCAGGGCCAAAGTCGCACAAC 1996
 QY 269 TyrAspAspCysHisAlaAspIleTyrGlnLysSerLeuLeuAsnLysArgTyrLeu 288
 Db 1996 ----- 1996
 QY 289 GlnLeuTyrProHisLeuLeuThrGluLysValAspTyrAspAsnProGlyLeuSer 308
 Db 1995 -----GACTATGACCAAGTTTCGGCAAC 1972
 QY 309 CysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLys----- 325
 Db 1971 TCCGGTTTCATGCAACCGCGCGAGCGAGCAACCTGTGCGCGCGCGCTTC 1912
 QY 326 -----ProTyr-----IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
 Db 1911 ACTCGTTGATCTCGTGTGACCTTCAGCCCTCAGCGCGCGCGAGGACAGGCGGAT 1852
 QY 341 TrpLysGlnPro-----PheProGlyPheGluAlaLeuArgValHisLeuAsn--- 356
 Db 1851 GCCTATCAGGATGCGCGTTTCTACTCGCGCTTCGATACCGTCCGCGACAGCTGCTCGG 1792
 QY 357 ArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsn 376
 Db 1791 CAGAACGACCTGACCCCTGCGCGAA-----GGCCATGTACTGACCCCTCGGCTACGACTGG 1738
 QY 377 GlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer 396
 Db 1737 CAGAGGACGAGATCAGCAGCAGCGAGCTTCAGCGCTCGACTCGCGCTGAACAAAGGC 1678
 QY 397 TrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVal 416
 Db 1677 TGGTTGCGCCAGTACCTCGCGCAGTACGGT----- 1648
 QY 417 LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu--- 435
 Db 1647 GCCAGGATGGCACTGAGCTGCGCGTGCAGCACACAGCAGTTCGGGCTGCACGAC 1588
 QY 436 -----GlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGln 453
 Db 1587 ACCGGCAGCGCGCGCTGG---GGCTACGCGGTGAGCGAGCGCGCTTC----- 1540
 QY 454 LysAlaSerIleArgTyrAsp-----LysAlaLeuIleAspArgGluAsnTyrTyr 470
 Db 1539 -----ACGTCATACGCGCAGCGCATTCAGGCGCGCGACTTCACGAAGTCTACTAC 1486
 QY 471 LysGln---ProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
 Db 1485 CCGGACTAGCGCAATCCGCGCTGCGCGAGAGCTTCGCGCAGCGCTGGAGGTCGGGCTG 1426

QY 490 SerGlyAsnTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
 Db 1425 AGCGGT----- 1420
 QY 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGly---LysHisValAlaThrAsnThr 528
 Db 1419 -----ACGCATGCTGGGGCACTGGCGCGTGAATGCC 1387
 QY 529 PheGlu-----ValGlyAsn----- 533
 Db 1386 TTCGTATCAACAGTCGACGACTGATCGCAACGATCGCGCTCCGGCTCCGGGCGCGCC 1327
 QY 534 -----LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
 Db 1326 TGGGGCGAGCGCAACATCATCAGCAAGCGCGCATCCCGCGCTCGAACTGGTCTCGGC 1267
 QY 550 TyrGluGlyAspArgTyrGlnTyrAsnLeu---AlaLeuTyr-----ArgAsnArg 565
 Db 1266 AGCCAGTGGTGGTGGACTGGACCGCAACCGCAACGCGCTTCCTCGACCCGCAAAACCGT 1207
 QY 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly-----ArgGly 580
 Db 1206 TCCGGCGGC-----GTCAACGACGCAACGAGCTCCCGCGCGCGCG 1165
 QY 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyVala 600
 Db 1164 CGCGGATGTTCAACCTGGAAGTGGCGCGCTATGATGATCGCGCAACAGGTGGCTGGGC 1108
 QY 601 AspPheTyrGlyAlaGluGlyLeuTyrPheLysProThrProArgTyrArgIleGly 620
 Db 1107 AGCGTGCAC---GCCGAAGCGCGCTATGATGATCGCGCAACAGGTGGCTGGGC 1051
 QY 621 -----ValSerGlyAspTyr-----ValArg 627
 Db 1050 GGTACGCGCCTCGACCTCGCGCGAGTACCGGCTGAACGAGAAATGGCGCTGCAG 991
 QY 628 GlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly---AsnArg 646
 Db 990 GGCGGATCGCAACCTGTTCGGT---GCCGACTAGAAACCGCTATGGCTACACCGAG 934
 QY 647 Pro 647
 Db 933 OCT 931
 RESULT 19
 US-09-221-017B-970
 ; Sequence 970, Application US/09221017B
 ; Patent No. 6444799
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSS, Bruce C.
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1120
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FORSTER
 ; STREET: 755 PAGE MILL ROAD
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FastSeq for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/221,017B
 ; FILING DATE: 23-DEC-1998
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP1182
 ; FILING DATE: 31-DEC-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP1546

```

; FILING DATE: 30-JAN-1998
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Montroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 970:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..4661
; US-09-221-017B-970

Alignment Scores:
Pred. No.: 9.44e-10 Length: 4661
Score: 196.00 Matches: 161
Percent Similarity: 35.95% Conservative: 105
Best Local Similarity: 21.76% Mismatches: 297
Query Match: 4.86% Indels: 178
DB: 4 Gaps: 37

US-09-936-377-2 (1-758) x US-09-221-017B-970 (1-4661)

QY 33 LeuGluThrValThrValValGlyLys-----SerArgProArg 45
Db 2282 TTGGAGGAGTCGTCGTACCGGTACCGGTACCGTTCGCTGGTCGATGCTCTGTG 2341
QY 46 AlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleLeSerGlyAspThr 65
Db 2342 GCACGGAAGTCCTTACCGGTAAAGACATAGCTCTTCTCGGCTCCTACTTCGAGGCC 2401
QY 66 LeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
Db 2402 TTATTGAG-----GGCTGAGTCGCTTTTTCGACTTCGGCCCAATCTGATGGCC 2452
QY 86 Ser-----GlnTyrglyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArg 103
Db 2453 TCTTTCATGAGCTGAACGGCCTTAGCAGTAAGTATATCTCATCTTATCGATGCTAAG 2512
QY 104 ArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAsp-----Phe 120
Db 2513 CGRTG-----TACGGCGATGTAGCGGTCAGGCCGATTTGAGTCGTATT 2557
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
Db 2558 TCTCTGTATCAGATC-----GAACGGATCGAACTGGTGAAA 2593
QY 141 GlyProValThrLeuLeuTyfyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 2594 GGTGCTTCGAGTTCGCTCTACCGATCCGATCCGATCGCGGGGTAACTAATCATGTGATC--- 2650
QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
Db 2651 -----ACAAAAAAGAAATACGAATCGAGTCAGTGCATATACGTCACATCGCAT 2698

```

```

Db      3565  CACCACTATGGTTCGCAC-----AATCTCTATCTCGCAATCGGATCTTAA 3612
Qy      537  nlysgluargserasnasnillegluLeuAlaLeuGlyTyrClucllyaspargtrpGlnTy 557
Db      3613  GCCACAGATGCGGATTAATATGCTTTGGGCTTGAGGTACAAATCAAGGCCCTATCTCGTT 3672
Qy      557  rasnLeuAlaLeuTyrArgasnArgPheGlyAsnTyrIleTyrAlaGlnThrLeu----- 575
Db      3673  CAGTGCACGGTTATGACATGACACTCGCAATCTGATCTCCTTTATGGATATACCGAC 3732
Qy      576  -----asnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLe 592
Db      3733  CTCACCCGAGCAGCAAGCTCGGGAATCAAGAAACCAAGCAGTATGCCACATAGGAAA 3792
Qy      592  uvalargtyrAsn-----GlnSe 598
Db      3793  AGCTCGAGCGCGCCCTTGATGTCCTATGTATGATGCTCTATCGGTTGGGTATCAAGTT 3852
Qy      598  rGlyAlaAspPheTyrGlyAlaGluGly-----GluIleTyrPheLysPr 613
Db      3853  AGGAGCCGGATACAGCCTCGTGGAGCTTAAGAAATCTCCAGAGCGGATGATGCTGGAAGG 3912
Qy      613  oThrProArgTyrArgIleGlyValSerGlyAspTyrValarg-----GlyArgLeuLy 631
Db      3913  AGCTGCAGTCATCGTGCCAAATGTCACCGCATGGGTTCACCTACTGGGTGAGTATAG 3972
Qy      631  sAsnLeuProSerLeuProGlyArg-----GluAspAlaTyr-----GlyAsnAr 646
Db      3973  ACTTGGCGTGAAGCCCTTTTCGGCGGTATTACAGAGCGAGCGTTACTACAAAGACGGCAATGC 4032
Qy      646  gProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPh 666
Db      4033  TCCG-----GACTATACCTGTGGCGACTCGCCCATCGCATCGCTTTCGC 4077
Qy      666  ehisLeuLysAlaSerLeuThrAsp-----ArgIleAspAlaAsnLeuAspTyrTy 683
Db      4078  TCATTTCCGCCACATCATCTCGTGAAGCGTCGGTATAGACAACTGTTGACTAGT 4137
Qy      683  r-----ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db      4138  GGATGATCGTCTATGGGTGTCAAT-----TATGCTACCGTAACGCCGGGA 4183

```

RESULT 20

```

; Sequence 1192, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1192
; LENGTH: 2391
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1192

```

```

Alignment Scores:
Pred. No.: 5,76e-10 Length: 2391
Score: 193.50 Matches: 170
Percent Similarity: 32.51% Conservative: 80
Best Local Similarity: 22.11% Mismatches: 234
Query Match: 4.79% Indels: 286
DB: Gaps: 37

```

```

US-09-936-377-2 (1-758) x US-09-252-991A-1192 (1-2391)
Qy      76  AlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaPro 95
Db      2377  GCGCTCGACGGTGAATGATCGACGGCTGCGGC-----GGCGGTGGCGCCTACGGCCA 2324
Qy      96  ValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGly 115
Db      2324  ----- 2324
Qy      116  AspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla-LeuSer----- 133
Db      2323  -----GCCGAGCTGTCGATGATGCCCTGTCCTATCGG 2291
Qy      134  -----GlnGlnValIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAs 151
Db      2290  CAACTGGAGAGCATCGACGTGTCGCGCGCGCTCGTTCGCTACGCGCCGCGAGAA 2231
Qy      151  nValAlaGlyLeuValaspValAlaAspGlyLysIleProGluLysMetProGluAsnGl 171
Db      2230  CGTGGCGGGGTGATCAACTTCGTGACCCCGGCGATTCGCGAGAGATTCTC----- 2182
Qy      171  YValSerGlyGluLeuGlyLeuArgLeu-----SerSerGlyAsnLeuGluLysLe 188
Db      2181  ----TCCGCGAAATCGCACCACTCGAGCATGCCGCCACCGCGCTGGAAGAAGCT 2126
Qy      188  uThr-----SerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHi 204
Db      2125  CAACCAAGCGCTTCCTCGCGGAGACCGCGCACCAACGCGCTGGC-----GTGGCGTGTCT 2072
Qy      204  sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLy 224
Db      2071  GTATTCGGGGTG-----AAGGGCCCGACTAC----- 2044
Qy      224  sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
Db      2043  -----CGCGACGGCAACAACGACCAACGATATCGACGACGTGCT 2006
Qy      244  pArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAl 264
Db      2005  GCTCAAGACCCANTTGGCGCTCACGACACGCGACCG-----TTGGCGGC 1961
Qy      264  aHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAs 284
Db      1960  CAATTCCTACTACTAC-----GACGCTTACGCGCATG----- 1927
Qy      284  nLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAs 304
Db      1926  -----CCGCGCGCGCTGACCCAGCGCAGTACGAC-----GACGA 1892
Qy      304  nProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGl 324
Db      1891  TCCCTTCAGTCGTA----- 1876
Qy      324  yLysProTyrIleAspLeuArgAsnLysArg----- 334
Db      1875  -CGGACTGGGACAAATTCGTGTCGCGCAAGGACTTCTCGTGAAGTACACCGCGCA 1817
Qy      335  -----TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPh 348
Db      1816  GGTGCGACGACTCACCCAGTTCAGGTGCTCACTACTACAGCGACATTTCCGCGCAG 1757
Qy      348  eGluAlaLeuArgValHisLeuAsnArg-----AsnAspTyrHisHisAs 363
Db      1756  CAGCATCGCGCGCGCACTCAGGACCATCCTCGTCCCGCGGCGACTACCTATGTGTT 1697
Qy      363  pGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIl 383
Db      1696  CGCGGTGAGCGCGGTCTCGCGGATCTCTTCCTCGCGCGCGCGACACC----- 1648
Qy      383  eGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGl 403
Db      1647  -----CAGGAGTGTGCGC-----ATCGGCTACCGTACTCTGAA 1616

```



```

Db 481 ---- CTTATTGAACGTGACCT----- 498
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db 499 ---- GAAAGCTTACCTCT----- 513
QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 514 ---- GAAACCTTATCCCGTCAAGCAGTGTCTGCTAGGTTTCATAT 558
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 559 GGGCGCTATT----- GACCAATAATTTGNAGCG 585
QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyr 260
Db 586 GCTGTGTGGCGATGAGAAATAA---TATATCCGTTTAAATGCCAACCGT----- 630
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIle----- 277
Db 631 ---- TCAGATCGAATAGTTATCAAGATGGTGTGCGAATACAGATTCATCAGCA 681
QY 278 TrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGlu 297
Db 682 TGGAAATAAGTGG----- 693
QY 298 GluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAla 317
Db 694 ---- AATGTTGATGAGCGCTTGGCTTTTACTCCAGATGAAACACAG 735
QY 318 HisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGlu--- 336
Db 736 ---- TGGTCGAAATCACGGGTGGGAAATCTGATGTT 768
QY 336 ---- 336
Db 769 GAATCGCTTTATGCTGGAGCTTCAATGATGATGTTCTCAGTTTGGCGGTGAAAGTTTAGGG 828
QY 337 LeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeu-----ArgValHis 354
Db 829 CTTTCGTTTTGAAAGAAATAATCACTGATGTGATCAAGAAATTTGAAGGGCAGGTGAAC 888
QY 355 LeuAsnArgAsnAspTyrHisHisAspGluLysAla----- 366
Db 889 TATAGCTATAACGACCATATTATGATACTTACCTTCGCGTACACCACTAGTTGAA 948
QY 367 ---- GlyAspAlaValGluAsnPhePheAsnAsn 376
Db 949 ATGAATCATGTTGTTGATGACCATGCTGATGTCGCAATGCAATGCGCAATGCAAGTGACACGC 1008
QY 377 GlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer 396
Db 1009 CGTACATTAACTCAGCTTTGGCAATGACTTCTGAA---TGGACAAATTTGAGTTTGATC 1065
QY 397 TrpGlyVal-----GlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGlu 414
Db 1066 ACAGGCATTGATTCTCAATTCAATAAACAATGCGGTGATGCTCATCGCCACATGCGCT 1125
QY 415 AlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyVal 434
Db 1126 TCAATGAATATATCCATATATCGTCAAGATATGCGTTTTCATCATACGGTGTGTTGGTGAG 1185
QY 435 GluGlnAlaAsnTyrAspAsnPheThr---LeuGluGlyGlyValArg----- 449
Db 1186 CTTGGTTATCAATGGAGCAATCTCAACAACTGGTCACAGGTGTCGCTTTAGATCGAGTC 1245
QY 450 ---ValGluLysGlnLysAla-----SerIleArgTyrAspLys 461
Db 1246 ACAGTTGAAGATGAACGGTGTGACTCTCAAGCAAAAGGCTTTAATACTAAGCTTGAAAG 1305
QY 462 AlaIleuLysAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481

```

```

Db 1306 ACTTTACCAGTGCATTTGTTCTGGGGAAAAATCAACGCTCGTGCACCTC----- 1353
QY 482 GlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeu 501
Db 1354 ---- GACTTCAAAAGC 1365
QY 502 SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla----- 519
Db 1366 TATATCGGTTGGGTTATGTAGAGCGTATGCGCTGATTTATTGGGAGCTCTTCTCGCCTGAG 1425
QY 520 HisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGlu 539
Db 1426 CATGTAATGCTGGCACAACATAATACCTTC-----AATGCTGTGAATCTCTGAG 1473
QY 540 ArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeu 559
Db 1474 AAAACA-----CTGCAGTTGGATCTTGGTTTCCAGCAGCAACATGGTCTTTAAATATA 1527
QY 560 ----AlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln----- 573
Db 1528 TGGGCTTCTGCTTATGCGAGGCTTAGTAGATGACTATATTTTGATGAATTTATCATGATCAT 1587
QY 574 ----ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSer 588
Db 1588 AGCCATCTGCATCCAAACGACATGCTGGCATGTGTAGTCATGATTTACGCTGTGTGCC 1647
QY 589 GluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu 608
Db 1648 AAA-----AATGTCGATGCAACCATCGCTGGTGCAGACGGGT 1686
QY 609 IleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGly 628
Db 1687 ATTGGATACCAATTCACTGACCGTATTGAGCGAGATTTGAGTGCATGTATGCTGGGGC 1746
QY 629 ArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPhe 648
Db 1747 ----AAAAAC----- 1752
QY 649 IleAlaGlnAspAspGlnAsnAlaProArgValPro-----AlaAlaArgLeuGlyPhe 666
Db 1753 ---ACCACAGATGACAAACCTCTACCTCAAAATTTTACCTTTAGAGGTCGTTTAAATATT 1809
QY 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPhe 686
Db 1810 CGTATGTTGGT-----GACAAATATAACTTAGGTCTGTATTGGCGAGCTGTTGCT 1860
QY 687 AlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMet----- 703
Db 1861 GAGCAAAACCGTGTGAGTCTACATCAAGTATATATCGTGGGATATGACCTAAACCAAGT 1920
QY 704 ----LeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyr 716
Db 1921 AAAGTTTTTCAACACTTCTTTAAACGGAATTAACCTTCGTAAGATATTGATGTC 1980
QY 717 GlyGluTyrAsnTyrTyrValLysAlaAspAsnLeu----- 728
Db 1981 TCT-----GTTGGCATTTGATAACGTTATTCGATAGACCTATACCGAACAC 2025
QY 729 LeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro-----GlnMet 746
Db 2026 TTAATAAAGCGGTAGTGCAGGCTTGGATTTGGAGGAGGAGGACAGTAAACAAATATC 2085
QY 747 GlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2086 GGAAGAAATTAAGTGGTTCGTATGATGATGAAGTTT 2121

```

RESULT 22

US-08-425-843-6

; Sequence 6, Application US/08425843

; Patent No. 6020154

; GENERAL INFORMATION:

; APPLICANT: Hansen, Eric J.

; APPLICANT: Cope, Leslie D.

APPLICANT: Jarosik, Gregory P.
 APPLICANT: Hanson, Mark S.
 TITLE OF INVENTION: H. Influenzae HxNB and HxuC Genes, Proteins
 TITLE OF INVENTION: H. Influenzae HxNB and HxuC Genes, Proteins
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: USA
 ZIP: 77210-4433

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION NUMBER: US/08/425,843
 FILING DATE: Concurrently Herewith
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Parker, David L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: AMCY.012/PAR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679
 TELEX: 79-0924

INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4651 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-425-843-6

Alignment Scores:
 Pred. No.: 3,53e-09 Length: 4651
 Score: 190.50 Matches: 163
 Percent Similarity: 35.66% Conservative: 123
 Best Local Similarity: 20.32% Mismatches: 293
 Query Match: 4.72% Indels: 225
 DB: 3 Gaps: 41

US-09-936-377-2 (1-758) x US-08-425-843-6 (1-4651)

Qy 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
 Db 742 CTTTCCCTTGCATTAACAACACC-----TTAGTGACGCAATGCACTAGCGCAATCC 795
 Qy 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
 Db 796 GTTGAATTAGACTCTCAACGTTATTGCGACACGAGATCCA-----AGTAGGTTT 846
 Qy 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
 Db 847 GCTTATAGCCAGAAAACATCTAAA-----GATAGTCTTCTTCTTACGACGCG 897
 Qy 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGly 90
 Db 898 ACTAGTGTGCGAGCGCTTAGAGACATTCCTCAATGTTGATGTTAGG-----GGCGGT 951
 Qy 91 Gly-----AlaSerAlaProValIleArgGlyGlnThrGlyArgArg-----IleLys 106
 Db 952 TCGAGACGATGCTCAAAAACCTTAATATCGAGGTTAGTGAATATCGTGTGTCGAA 1011
 Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
 Db 1012 GTCAATTGATGGCGTGAGACAAAATTTTGTATGACACATAGAGT-----TCTTAT 1062
 Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeu 146

Db 1063 TTTCTTCCATGTCACTTATTCAAGAAATTAAGTAATCAAGAGACCAAGTAGCTCCTTA 1122
 Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166
 Db 1123 TGGGTAGCGGTGCTTGGGTGGTGTGTGCAATGCGTACGCCAAATGCTTTAGACTTA 1182
 Qy 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly----- 183
 Db 1183 TTGAATAATAT-----GACAAATTCGAGTAAATTCGCAAGGTATCAACT 1233
 Qy 184 -----AsnLeuGluLys----- 187
 Db 1234 GCTAATAATTTATCGAAAAGGATGCTTCTGTATTGCGCAAAATGACAAATTCGATGT 1293
 Qy 188 -----LeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeu 203
 Db 1294 CTTATTAGTGTCTTCTAATAATAGCGGATAATTACGACTGGTAAAGGACACAGCTG 1353
 Qy 204 HisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg----- 221
 Db 1354 AATAATACAGCC--TATAAACAGTTTGGGGCTTAGCAAAATTTGGTTGGCAATCAAT 1410
 Qy 222 -----AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
 Db 1411 GATCGAACCGGTGTGGAATTTATCCACCGCAAACTCGTTTAA--AACAACAGCACCA 1468
 Qy 240 AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
 Db 1469 GCATAATGAGGTGGAAACAGCACTTACC--AATGAAAAAATTTATAGATCAATCAACGAG 1527
 Qy 260 Tyr-----GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla 274
 Db 1528 TTCCACGGCTCAACACACGCTTTACACAGAGACAAACCATCATCAACATCAACGCG 1587
 Qy 275 AspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLysTyrProHisLeu 294
 Db 1588 TTTTAC-----TCTAAAGTGAAACACGCTTTGGCAGTGTC-----AGTTAT 1629
 Qy 295 LeuThrGluGluAspVal-----AspTyr-----Asp 303
 Db 1630 TTAACGTGATCAACAAATTCCTGATCAAGACGCGTATTAACTATTATTAAACGCGAGAT 1689
 Qy 304 AsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHis 323
 Db 1690 AATCCTTATCTA-----AATACGATATGCGACTGCTGATTAAC 1725
 Qy 324 GlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluThrLysGln 343
 Db 1726 AATAAAACTATTGAGAAAGACAGCGTAAAGTCAGTGGTGTGAAAGATCAGACTAAATG 1785
 Qy 344 ProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
 Db 1786 ACTACCGGAGGT-----ATAAATTACGTAACTCTTCGAAATTAATCTCTAC 1830
 Qy 359 -----AspTyrHisHisAspGlyLysAlaGlyAspAlaValGlu 371
 Db 1831 ATTTCCTTGTGTTATGGGTGGATTATATCGCAGATAAAATCCGTACCGAA----- 1881
 Qy 372 AsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProLeuGly 391
 Db 1882 -----CGAGGCAACAACTAATGATGCGAATTTTCGAGCGGACCCCTATAAT 1929
 Qy 392 ArgLeuLysGlySerTyrProGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla 411
 Db 1930 GCGAATTCAAACATACACGCGGT--TATTTA----- 1959
 Qy 412 ThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhe 431
 Db 1960 -----ATCGCCCATATTCCACTA 1977
 Qy 432 PheGlyValGluGlnAlaAsnTrpAspAsnThrLeuGluGlyValArgValGlu 451

Db 2848 ATGGTGGCTATGATCGCGCGTGGTCTTCTGTCGATGGTGTTCGTCACGGGAACG 2789
 QY 113 GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
 Db 2788 GATACCGACACCTG- 2747
 QY 133 SerGlnGlnValGluLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnVal 152
 Db 2746 ATCAAGCGTGTGAGATTGTCGTGGACCTTCAGCAATTACTGTATGGCAGTGGCGCGCTG 2687
 QY 153 AlaGlyLeuVal-AspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
 Db 2686 CGTGAGTGTATCTCTACGATACGCTCGATGCAAAA- 2636
 QY 171 GlyValSerGlyLeuGlyLeuArgLeu--SerSerGlyAsnLeuGlyLysLeuThr 189
 Db 2635 GACAAAGAGT- 2597
 QY 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu- 208
 Db 2596 ACGGGGACCATAGCTGGGANTAGCGCGAGCGGCTTTGGCGCAACTGAAAAATCTGGAT 2537
 QY 209 - 216
 Db 2536 GGTATTGGCTGGTCCAGTCGATCGGATCGGGTGTATTACGCCAGCAATGTTGAAC 2477
 QY 217 ValProArg- 219
 Db 2476 GCGCGAATGACGAGTCCATTAAATACATGCTGGCGAAAGGACCTGGCAAAATGATTCA 2417
 QY 220 - 229
 Db 2416 GCCAGTCTCTGAGCGGTTTGTGCGTTACTACAAACAGCGCGGTGAACCAAAAAT 2357
 QY 230 ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrPArgLysArgPheTyr 249
 Db 2356 CCGCAGACCGTTGGGCTTCTGAAGACGACCAACCGATGTTGATCGTCAACAATTCAA 2297
 QY 250 ArgArgThrTyrSerAspArgaspGlnTyrGlyLeuProAlaHisSerHisGluTyr 269
 Db 2296 CGC- 2249
 QY 270 AspAspCysHisAlaAspIleLeuTyrGlnLysSerLeuIleAsnLysArgTyrLeuGln 289
 Db 2248 TTAATGCGAGTCAAAAATTATTTGGTGGAGTCCGTTATTAAATGCGCAA- 2198
 QY 290 LeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys 309
 Db 2197 - 2180
 QY 310 GlyPheHisAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrPileAsp 329
 Db 2179 GAGTATCGTGAACA-GATACAAAGAGCGCGCT- 2127
 QY 330 Leu-ArgAsnLysArgTyrGluLeuArgAlaGluTyrPlyGlnProPheProGlyPhe 348
 Db 2126 CTTTGGCGACAGTTTCGCTCTCACTTACTGACATATGG- 2088
 QY 349 GluAlaLeuArgValHisLeuAsnArg-AsnAspTyrHisHisAspGluLysAlaGlyAs 368
 Db 2087 - 2052
 QY 368 PAlaValGluAsnPhe- 382
 Db 2051 CGCGACGCGGCTTCCCGCAAGCAAAAATCGATTTTGTAGCTCGCGTGGCTACAGATGA 1992
 QY 382 GileGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLe 402
 Db 1991 GATCACCTTACGGGATCTGCGGAT- 1967
 QY 402 uGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeu 422
 Db 1967 - 1967

QY 422 uAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPh 442
 Db 1967 - 1967
 QY 442 eThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAl 462
 Db 1966 -ACCTCTGTTGGCGAACCGCTATGACAGTTATCGCGTAGC- 1925
 QY 462 aLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyValaHisArgI 482
 Db 1924 -AGTGACGGTTTACAAA- 1890
 QY 482 nThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSe 502
 Db 1889 GTCATCTCGTGC- 1845
 QY 502 rLeuThrAlaSerHisGlnGlu-ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 Db 1844 GTTATTGGCTCATATGCCCGAGCATTCGCGCGCCGAGATGGCGGAATGTATAACGA 1785
 QY 520 sGlyLysHisValAlaThrAsnThrPheGlu-ValGlyAsnLysHisLe 536
 Db 1784 TTCTAGCACTTCTCGATTGGTGGTCTTATACCAACTATTGGTGCACAAACCCGAAC 1725
 QY 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly- 549
 Db 1724 ACGTCGGAACACTAACGAAACTCAGGAGTACGGTTTTGGCTGCGT- 1665
 QY 550 -TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
 Db 1664 GTTGTCCATGATGCTCTGGAATTTAAAGCCAGCTACTTTGATACCAAGCGAAGATTA 1605
 QY 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGl 589
 Db 1604 CATC- 1578
 QY 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla- 607
 Db 1577 GCGCAGCACTATGTGTATTAACCTCGGAACCGCAAAATCTGGGCTG- 1519
 QY 607 yGluIleTyr- 618
 Db 1518 CGAATATACCACTACTCTGTAGCTTGTATGGCTTATACCGTACCGCGGCAAG 1459
 QY 618 gIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGl 638
 Db 1458 A- 1411
 QY 638 yArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProAr 658
 Db 1410 CCAGCACTCTGAATATTCGATCGCTCACA-GTGGCTTCTCTGTGGGTGGTGTACG 1352
 QY 658 gValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThr 673
 Db 1351 TTTCCGATCGCTCAA- 1315

RESULT 24

US-08-487-890A-4
 ; Sequence 4, Application US/08487890A
 ; Patent No. 5708149
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney
 STREET: 6th Floor, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,890A
 FILING DATE: 07-JUN-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/175,116
 FILING DATE: 29-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/148,968
 FILING DATE: 08-NOV-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I.
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5099 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Join(160..2121, 2152..4890)
 US-08-487-890A-4

Alignment Scores:
 Pred. No.: 4,536-08 Length: 5099
 Score: 180.50 Matches: 197
 Percent Similarity: 33.30% Conservative: 139
 Best Local Similarity: 19.52% Mismatches: 323
 Query Match: 4.47% Indels: 351
 DB: 1 Gaps: 48

US-09-936-377-2 (1-758) x US-08-487-890A-4 (1-5099)

QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
 Db 2155 ACTAAAAACCCCTATTTCGCCCTAAGTATTATTCTTGCTTTTATTTCATCTATGTA 2214
 QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
 Db 2215 AAAGCAGAACTCAAGTATAAAGATACAAAGAGCTATATCATCTGAAAGTGACACT 2274
 QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
 Db 2275 CAAAGTACAGAGATTTCAGAACTAGTAATACTCTCAGTCTGCGAATAAATAAGAGAT 2334
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
 Db 2335 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATCAAACTAGT 2382
 QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
 Db 2383 GAAAGTATCAGCGAGAACAGATTAATAATATTCTGTGATCTAACACGCTATGATCCAGC 2442
 QY 83 IleHisAlaSerGlnThrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
 Db 2443 ATTCAGTTGTAGAACAGGCCGTGGTGCAAGTTCTGGATATTCTATTCTGCGTATGGAC 2502

QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
 Db 2503 AGAAATAGATT----- 2514
 QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly 141
 Db 2515 -----GCTTATTAGTAGATTGGTTACCTCAACAGCAATCTTATGTAGTCCAAAGC 2565
 QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
 Db 2566 CCTTTAGTCTGCTGCTCAGGATATTCTGGCACTGGTGCAATTATGAATAATGAATATGAA 2625
 QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179
 Db 2626 AATGTAAAGCCGCTCGAA---ATAAGCAAGGGGGAGTTCTTCTGAGTAGTGGTAATGGA 2682
 QY 180 LeuSerSerGlyAsnLeuLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198
 Db 2683 GCCTAGCTGGTTCTGTACATTTCAAAGCAAAATCAGCAGCCGATATCTTAGAAGGAGAC 2742
 QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210
 Db 2743 AATCATGGGAATTCAAACTAAATAATGCTTATCAAGCAAAATAAAGGCTTTACCCAT 2802
 QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
 Db 2803 TCTTTAGCTGTAGCTGGAACCAAGGGGATTTGACGGGTCCGCAATTTATCTCAACGA 2862
 QY 222 AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala 240
 Db 2863 AAT-----TCAATTGAAACCAAGTCCCATAAAGATGCA 2895
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260
 Db 2896 TTAAGAGCGGTACAAAGT-----TATCATCGATTATATGCCCAACCCAGAGGATCAA--- 2946
 QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleThrGlnLys 280
 Db 2947 -----TCTGCATACTTTGTGATGCAAGATGAGTGTCCAAAGCCAGATGATTATACAGT 3000
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
 Db 3001 TGTTTACCTTTGCCAAACGACCTGCGATTTTATCTCTCCAAAGAGAAACCGTAAGCGTT 3060
 QY 293 -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy 309
 Db 3061 TCAGATTATACGGGGCTAACCGTATCAAACTCAATCAATCAATGAATGAAGCCAGTCT 3120
 QY 309 SGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProThrPleAs 329
 Db 3121 TGGTTTTTAAGAGGGGGGTAT-CATTTTCTGAACCAACATATATATTTGGTGTATTTTGA 3179
 QY 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349
 Db 3180 ATTCACACACAAATAATTTGATATCCGT-----GATATGACATTTCCCGCTTAT-- 3228
 QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlyLeuAlaGlyAspAl 369
 Db 3229 -----TTAAGATCAACAGAAAAACGGGATGATAG 3257
 QY 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
 Db 3258 CAGTGGCTTTTTTATCCAAAGCAAGATTTATGTCATATCAACGCTATTGAG----- 3309
 QY 387 sdlnProIleGlyArg----- 393
 Db 3310 -----GATGCCCGAGGGCTTAACATGCAAGTGGGCTTTATTTTCGATGAACACCATAG 3362
 QY 393 uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
 Db 3363 AAAACAGCGGTAGGTATTGAATATATTACGAAAAATAAGAACAAAGCGGCGCATTTCA 3422
 QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis--- 428

Db 3423 CAAAGCAGTGTAAAGTCTAATCAACAAACATCATCTTACATGACATTTATATGCAACATAC 3482
QY 429 -----TyrValArgGlyArgLeuLysAsnLe 633
Db 3483 GCATTGCGAGTCTTTATCTCTAATCAAGTAAGAAATTGCGCCCAACACAGTGTATAACCTTA 3542
QY 429 rSerPheGlyValGlu----- 435
Db 3543 TTCACTACTCATCTGTATGAAGATCTTTATAAGAAAAACATAATATGTTGCAATTGAA 3602
QY 436 -----GlnAlaAsnTrp----- 439
Db 3603 TTTAGAGAAAAAATTTCAACAAAATTTGCTTACTCATCAAAATTTCTTCAATCTTGGTTT 3662
QY 440 AsnAsnPheThr-----LeuGluGlyValArgValGI 451
Db 3663 TGATGACTTTACTTCAGCGCTTCAGCAATAAGATTTTAACTGACGCTGTTACCGCTAC 3722
QY 451 uysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLy 471
Db 3723 GGCAGAGATTTTCA-----GAGAAAGCTAATGAAACAGAAATGTTTCAA 3773
QY 471 sGln-----ProLeuProAspLeuGly----- 478
Db 3774 AAAACAACTTACTTATACCAAAACCAACAGTAGTGTTTGTGTACAGATCATTTGTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAGGTAACTCTCTAATTAACAGAGACTGTAAAGTGGCGTTAATTAAGGGAAGAAA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu----- 503
Db 3894 TTATTATTTCGACGACGCAATTAATATGCGATTAGCGGAATACGTTGATTTAGGTTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu----- 509
Db 3954 TATTCGGTATGAGTATCTCGCACAAAGCTAATGATCACTATTAGTGTGGTAATTT 4013
QY 509 ----- 509
Db 4014 TAAAAATTTCTCTTGAATACTGGTATTTGTCATAAAACCAACGGAATGGCTTGATCTTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db 4074 TTATCGCCTTCTACTCGGATTTAGAAATCTAGTTTGTCTGAATGTAT-----GGTGG 4127
QY 522 shisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGGTGGCAATAATAGCGAGGTTTATGTAGGTAAATTTAAGCCTGAAACATCTCG 4187
QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAACCAAGATTTGGTCTCGCTCTAAAGGGATTTTGTGTAATATGAGATCAGTCAITTT 4247
QY 562 rArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsn-----AspGI 578
Db 4248 TAGTAATGCTTATCGAAATCTTATCGCTTTCGTAAGAACTTAATAAAGTGAAGTGG 4307
QY 578 yArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn---GI 597
Db 4308 AAAAGGCCAATATGATATATCAATATCAACAAATGCAAAATAGTTGGCGTAAATATAAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db 4368 TGCGCAATTAGATTTTAATGTTTATGGAACGATTTCCCTACGCTTGTGTATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCTTATACCGAGTAAAAAGTTAAAGATCAAAAATCAATGCTGCTGGTTCGCTCGGTAAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 612 ----- 624

4488 CAGTTATTATTGATGATCCATTGAGCCAGC-----CGTTATATCATTTGGTTTAGGCTATCA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCAAGTATATCTTGGGGAATTAATCAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTTGCTAGGAAACAGT---GCATTGGGTAAACAAT-----TCAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAAATCAACAAAGAAATTTACTCGGCA-----TGGCAATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATCTATCGGTTTATTACATGCGCAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 4723 -----AATATTATGCTTCGATTAGGATATATAATTTATTTCAA 4760
QY 713 nThrArgTyrGlyGluTyrAsnTyrTyrValLys---AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGCTATGTTACTTGGAGCGGTGCTCAACAGCAGCAGGTCGGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGVAGCTATACTCTGTCACGAGCATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 25
US-08-478-435-4
; Sequence 4, Application US/08478435
; Patent No. 5922323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160..2121, 2152..4890)
; US-08-478-435-4

Alignment Scores:
Pred. No.: 4 53e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 2 Gaps: 48

US-09-936-377-2 (1-758) x US-08-478-435-4 (1-5099)

QY 5 ThrLeuLysProfile-----ValLeuSerIleLeuLeuIleAenThrProLeu 20
Db 2155 ACTAAACCAACCTATTTCGCCCTAAGTATTATTTCTTGCTTTATTTATTCATGCTATGTA 2214
QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly 32
Db 2215 AAAGCAGAACTCAAAAGTATAAAGATACAAAGAAAGCTATATCATCTGAAGTGGACACT 2274
QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 2275 CAAAGTACAGAGATTCAAAATTAGAACTATCTCAGTCTCGACGAAAGAAATAAGAGAT 2334
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle---Ser 62
Db 2335 CGTAAAGATAATCAAGTA-----ACTGGACTTGGCAAAATTATCAAACTAGT 2382
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2383 GAAAGTATCAGCGGAGAAACAGATTATAATATTTCGTGATCTAACACGCTATGATCCAGGC 2442
QY 83 IleHisAlaSerGlnThrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
Db 2443 ATTTCAAGTTGTAGAACAGCGCGTGTGCAAGTTCGGATATTCTATTCTGGTATGGAC 2502
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 2503 AGAAATAGAGTT----- 2514
QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGly 141
Db 2515 -----GCTTTATTAGTAGTGGTTTACCTCAAAACGCAATCTTATGTAGTGCAAGC 2565
QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 2566 CTTTAGTTGTCGTCGTTCAGATATTCGCACTGGTGGCAATTAAAGAAATTGAATATGAA 2625
QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179
Db 2626 AATGTAAGGCCGTCGAA---ATAAGCAAGGGGGGAGTCTCTCAGTATCGTATGGA 2682
QY 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198
Db 2683 GCACCTAGCTGGTTCGTAAACATTTCAAGCAAAATCAGCAGCGCATATCTTAGAAGGAGAC 2742

```

```

QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyArg----- 210
Db 2743 AAATCATGGGAATTCAAACTAAATAATGCTTATTACGACAAAATAAAGGCTTTACCCAT 2802
QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
Db 2803 TCTTTAGCTGTAGCTGGAAAACAAAGGGGATTTGACGGGCTCGCCATTTATATCAACGA 2862
QY 222 AsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArg---Ala 240
Db 2863 AAT-----TCAATTCAAAACCCCAAGTCCATAAAGATGCA 2895
QY 241 ValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260
Db 2896 TTTAAAGGGCGTACAAAGT-----TATCATCGATTAAATCGCCAAACCCAGAGGATCAA-- 2946
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTrpGlnLys 280
Db 2947 -----TCTGCATACATTTGTGATGCAAGATGAGTGTCCAAAGCCAGATGATTATAACAGT 3000
QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
Db 3001 TGTTTACCTTTCGCCAAACGACCTGGATTTCCTCCCAAGAGAAACCGTAAGCGTT 3060
QY 293 -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy 309
Db 3061 TCAGATTATACGGGGCTAACCGTATCAAACTTAATCAATGAATATATGAAGCCAGTCT 3120
QY 309 sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAs 329
Db 3121 TGGTTTTTAAGAGAGGGGTAT-CATTTTCTGAACCAACATTAATTATGGTGTTATTTTGA 3179
QY 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheG 349
Db 3180 ATTCAACACAAACAAATTTGATATCGT-----GATATGACATTTCGCCCTAT-- 3228
QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlyLysAlaGlyAspAl 369
Db 3229 -----TTAAGATCAACAGAAAAACCGGATGATAG 3257
QY 369 aValGluAsnPheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
Db 3258 CAGTGCTCTTTTTCACAAAGCAAGATTATGTTGTCATATCAACGTATTGAG----- 3309
QY 387 sGlnProIleGlyArg-----Le 393
Db 3310 -----GATGCCCGAGGGTTAACTATGCAAGTGGGCTTTATTTGATGAACACCATAG 3362
QY 393 uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
Db 3363 AAAACACAGCGTGTAGTATTGAATATATTACGAAAAATAAGAACAAAGCGGCGCATTTGA 3422
QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
Db 3423 CAAGCAGTGTAAAGTGTCTTAATCAACAAACATCATACTACTGACGTTATATGCAACATAC 3482
QY 429 -----TY 429
Db 3483 GCATTGCACTTTTATCTTAATCCAAGTAAGATTGCCGCCCAACACGTGATAAACCTTA 3542
QY 429 rSerPhePheGlyValGlu----- 435
Db 3543 TTCTACTATCATCTGTATAGAAATGTTTATAAGAAAAACATAATATGTTGCAATTGAA 3602
QY 436 -----GlnAlaAsnTrp----- 439
Db 3603 TTTAGAGAAAAAATTCAACAAATTTGGCTTACTCATCAAAATTTGCTTCAATCTTGGTTT 3662
QY 440 -AspAsnPheThr-----LeuGluGlyGlyValArgValG 451
Db 3663 TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTATTAACTCGACGCTGTTACCGCTAC 3722

```



```
QY 451 ulysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrly 471
Db 3723 GGCAGAGAGTATTTC-----GAGAAAGCTAATGAACAAGAAAGAAATGGTTACAA 3773
QY 471 sGln-----ProLeuProAspLeuGly----- 478
Db 3774 AAAACAACCTTACTTATACCAAAACACAGTAGGTTTGTAGTACAAGATCATTTGTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGGTAACTCTCTAATACAGAGACTCTAAAGTCGCGTTAATTAAGGGGAAAA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu----- 503
Db 3894 TTATTATTTCGCAGCAGCAATAATATGCGATTAGGCAATACGTTGATTAGGTTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu----- 509
Db 3954 TATTCGGTATGAGTATCTCGCACAAAGCTAATGAATCACTATTAGTGTGGTAAATT 4013
QY 509 ----- 509
Db 4014 TAAAAATTTCTTGGAACTACTGGTATTGTCTATAAAACCAACGGAATGGCTTGATCTTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db 4074 TTATCGCTTTCTACTGGATTAGAAATCTCTAGTTTGTGTAATGAT-----GTTG 4127
QY 522 sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGGTGCATAATAAGCAGAGTTTATAGGTAAATTTAAGCCTGAAACAICTCG 4187
QY 542 nAnlIleGluLeuAlaLeuGlyTyrGluGlyAspArgGfrGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAACCAAGAGTTTGGTCTCGCTCTAAAGGGGATTTTGGTAATATTGAGATCATCTATT 4247
QY 562 xArgAsnArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsn-----AspG1 578
Db 4248 TAGTAATGTTATCGAAATCTTATCGCTTTGTGGAAGACTTAATAAAATGGAATGG 4307
QY 578 yArgGlyProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsn---G1 597
Db 4308 AAAGGCCAATTATGGATATCATATGACAAATGCAAAATAGTATGTTGGCGTAAATTAAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db 4368 TGGCAATTAGATTTTAAATGTTTATGGAACGTAATCCCTACGTTGGTATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCTTATAACGAGTAAAGTTAAAGATCAAAATCAATGCTGGTTTGGCCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTGATGCCATTACGCCAGC---CGTTATCATTTGGTTTGGCTATGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAGTAATACTTGGGGAATTAATACAAATGTTTACTCAATCAAAACCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTCTAGGAAACAGT---GCATTGGGTAAACAAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAGAAACTTACTCGGCA-----TGGCATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrGValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGATCGGTTTATTATCGCGGAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
```

```
Db 4723 ----- 4760
QY 713 nThrArgTyrGlyGluTyrAsnTyrTyrVallys---AlaAspAsnLeuLeuAsn--- 730
Db 4761 CTATCGCTATGTTACTTGGAGCGGTGGTCAACAGACACAGAGTGGCGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATACTCGTACGCAAGCATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnVallysPhe 758
Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 26
US-08-337-483-4
; Sequence 4, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160..2121, 2152..4890)
US-08-337-483-4

Alignment Scores:
Pred. No.: 4,53e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 2 Gaps: 48

US-09-936-377-2 (1-758) x US-08-337-483-4 (1-5099)
```

Qy	5	ThrLeuLysProIle-----ValleuSerIleLeuLeuIleAsnThrProleu	20
Db	2155	ACTAAAAACCCCTATTTTCGCCTAAGTATATATTTCTGTCTTTTAATTCATGCTATGTA	2214
Qy	21	LeuAlaGlnAlaHis-----GluthrGluGlnSerValGly-----	32
Db	2215	AAAGCAGAAACTCAAAAGTATAAAAGATACAAAGAGCTATATCATCTGAAGTGCACACT	2274
Qy	33	-----LeuGluThrValThrValValGly---LysSerArgPro	44
Db	2275	CAAAAGTACAGAGAGATTCAGAAATTAGAAACTATCTCAGTCACTGCAGCAAAAATAAGAGAT	2334
Qy	45	ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser	62
Db	2335	CGTAAAGATATGAAGTA-----ACTGCACTTGCCAAAATTTATCAAACTAGT	2382
Qy	63	GlyAspThrLeuArgGlnGlyAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly	82
Db	2383	GAAAGTATCAGCCGAGAACCAAGTATTAATATTCGTGATCTAAACACGCTATGATCCAGGC	2442
Qy	83	IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal-----IleArgGlyGlnThr	101
Db	2443	ATTTCAAGTTGTAGAACCAAGGCGGTGTCAGAGTTCTCGATATTTCTATTCGTGGTATGCAC	2502
Qy	102	GlyArgAlaGlyLeuValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer	121
Db	2503	AGAAATAGAGTT-----	2514
Qy	122	ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly	141
Db	2515	-----GCTTTATTAGTAGATGGTTTACCTCAAACGCAATCTTATGATGTCACAAAGC	2565
Qy	142	ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla	159
Db	2566	CCTTTAGTGTGCTGTCAGGATATTTTCGGCACTGCTGCAATTAATGAATGAATGAATGA	2625
Qy	160	AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg	179
Db	2626	AATGTAAGGCGCTCGAA--A7AAGCAAGGGGGGAGTCTTCTGTAGTATGGTAATGGA	2682
Qy	180	LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly---	198
Db	2683	GCACCTAGCTGCTTCTGTAACATTTCAAAGCAAAATCAGCAGCCGATATCTTAGAAGGAC	2742
Qy	199	LysAsnPheValLeuHisThrGluGlyLeuTyrArg-----	210
Db	2743	AAATCATGGGAATTCAAACTAAAAATGCTTATTTCAGCAAAAATAAAGGCTTTACCCAT	2802
Qy	211	-----LysSerGlyAspTyr-----AlaValProArgTyrArg	221
Db	2803	TCITTAGCTGAGTGGAAAAACAAGGGGATTTCACGGGGTCGCCATTATATCTCAACGA	2862
Qy	222	AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala	240
Db	2863	AAT-----TCAATTGAAACCCCAAGTCCCATAAAGATGCA	2895
Qy	241	ValleuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr	260
Db	2896	TTAAAGGCGCTCAAAAGT-----TATCATCGAATTAATCGCCAAAACAGAGAGTCAA---	2946
Qy	261	GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTrpGlnly	280
Db	2947	-----TCTGCATACCTTTGATCCAGATGAGTGTCCAAAGCCAGATGATTATACAGT	3000
Qy	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro-----	292
Db	3001	TGTTTACCTTTTCGCAAAACGACCTGCGATTTTATCTCCCAAGAGAAACCGTAAACGTT	3060
Qy	293	-----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy	309
Db	3061	TCAGATTATACGGGGGTAAACCGTATCAACCTTCAATCCAAATGAATATGAAGCCAGTCT	3120

QY	309	sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpLeuAs	329
DB	3121	TGGTTTAAAGAGAGGGGTAT-CATTTTTCTGAACACATATATTGGTGGTATTTTGA	3179
QY	329	pleuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheG1	349
DB	3180	ATTCAACAACAAAAATTGATATCCGT-----GATATGACATTTCCCGCTTAT--	3228
QY	349	uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAl	369
DB	3229	-----TTAAGATCAACAGAAAACCGGATGATAG	3257
QY	369	aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi	387
DB	3258	CAGTGGCTCTTTTATCCAAAGCAAGATTATGGTGCATATCAACGTATTGAG-----	3309
QY	387	sGlnProIleGlyArg-----	393
DB	3310	-----GATGCCGCGAGGGTTACTATGCAATGGGCTTTATTCGATGAACACCATAG	3362
QY	393	uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe	413
DB	3363	AAACACGGCTGTAGTATTGAATATATTACGAAATTAAGAACAAAGCGGCATCATGCA	3422
QY	413	rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis--	428
DB	3423	CAAAAGCAGTGTAAAGTGCTAATCAACAAAAACATCATCTTGACAGTTATATGCAACATAC	3482
QY	429	-----GlnAlaAsnTrp-----Ty	429
DB	3483	GCATTGCAGTCTTTATCCTTAATCCAAGTAGAAGATGCCGCCCAACACGTGATAAACCTTA	3542
QY	429	rSerPhePheGlyValGlu-----	435
DB	3543	TTCATACTATCATCTGATAGAAATGTTTATAAGAAACACATAATATGTTGCAATTGAA	3602
QY	436	-----GlnAlaAsnTrp-----	439
DB	3603	TTTAGAGAAAAAATTCAACAAAATTGGCTTACTCATCAAAATGTCTCAATCTTGTTT	3662
QY	440	AspAsnPheThr-----LeuGluGlyGlyValArgValG1	451
DB	3663	TGATGACTTTACTTTCAGCGCTTCAGCATAAAGATTATTAACTCGACGTGTTACCGCTAC	3722
QY	451	uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTrpIly	471
DB	3723	GGCAACAGAGTATTCA-----GAGAAAGCTAATGAACAGAGAAATGGTTACAA	3773
QY	471	sGln-----ProLeuProAspLeuGly-----	478
DB	3774	AAAAACAACCTTACTTATACCACAAACCAACAGTAGGTTTTGTAGTACAAGATCATTTGTA	3833
QY	479	-----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs	492
DB	3834	TTATAAGGTAACCTCCTTAATTACAGAGACTGTAAGTGCGGTTAATTAAAGGAAAAA	3893
QY	492	nTrpTyrPheTrpProGlnHisLysLeuSerLeu-----	503
DB	3894	TTATTATTCGACGACGCAATAATATGCATTAGGAAATACGTTGATTAGGTTTAGG	3953
QY	504	-----ThrAlaSerHisGlnGlu-----	509
DB	3954	TATTTCGTATGACGTATCTCGCAACAAAGCTAATGAATCACTATTAGTTGGTAAATT	4013
QY	509	-----	509
DB	4014	TAAAAATTTCTCTTGGATACTGGTATTGTGTCATAAAACCAACGGAATGGCTTGATCTTTC	4073
QY	510	-----ArgLeuProSerThrGlnGlnLeuTyrAlaHisGlyLy	522
DB	4074	TTATCGCCTTCTACTGGATTATTAGAAATCTTAGTTTTCGTGAATGTAT-----GGTTG	4127
QY	522	sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs	542

2383	DB	GAAGTATCAGCCGAGCAACAGTATTAAATATTCTGGTACTTAACACGCTATGATCCAGGC	2444
83	QY	ILEHISALASERGLNTRYGLYGLYGLYALASERALAPROVAL---ILEARGGLYGLNTHR	101
2443	DB	ATTTCAAGTTGTAGAACAAAGCCGTGGTGGCAAGTTCTGGATATTCTATTCGTGTATGGAC	2502
102	QY	GLYARGARGILELYSVALLEUASNHISGLYGLUTHRGLYASPMETALAAAPPHESER	121
2503	DB	AGAAATAGAGTT-----	2514
122	QY	PROASPHISALAIEMETVALASPTHRALALEUSERGLNGLNVALGLUILEULEUARGGLY	141
2515	DB	-----GCTTTATTAGTAGAGTTTACCTCAAAGCAATCTTATGTAGTGCAGCAAGC	2565
142	QY	PROVALTHRLEULEU-----TYRSESRGLYASNVALALAGLYLEUVALASPVALLA	159
2566	DB	CCITTAGTTCGTTCAGATATTTTCGCACATTCCTGGCACTGGTGCAAATTAATGAAATGAAATGAA	2625
160	QY	ASPGLYLYSILEPROGLIULYSMETPROGLUASNGLYVALISERGLYGLULEUGLYLEUARG	179
2626	DB	AATGTAAGCCCGTCGAA--ATAAGCAAGGGGGGAGTCTTCTGAGTATGTAATGGA	2682
180	QY	LEUSERSERGLYASNLEUGLULYSLEUTHRSERGLYGLYLEASNILEGLYLEUGLY---	198
2683	DB	GCACGTAGCTGTTCTGTAAACATTTCAAAGCAAAATCAGCAGCCGATATCTTTAGAAGCAGAC	2742
199	QY	LYEASNPHVALLEUHLISTHRGLUGLYLEUTHYARG-----	210
2743	DB	AAATCATCGGGGAATTCAACTAAAAATGCTTATTCAGACAAAATAAAGCGCTTATCCCAT	2802
211	QY	-----LYSESRGLYASPTYR-----ALAVALPROARGTYRARG	221
2803	DB	TCITTTAGCTGTAGTCGAAAAACAAGGGGGATTTGACGGGGTCGCCATTTTACTCAACGA	2862
222	QY	ASNLEULYSARGLEUPROASPSERPROARGPHEALASNGLYGLNHISARG--ALA	240
2863	DB	AAT-----TCAATTGAAACCCCAAGTCCCATTAAGATGCA	2895
241	QY	VALLEUGLYTRPARGLYSARGPHETYRARGTHRTRYRSERASPARGARASPSGLNTRYR	260
2896	DB	TTAAAGGCGTACAAAGT-----TATCATCGATTAAATCGCCAAACAGAGGATCAA--	2946
261	QY	GLYLEUPROALAHISERHISGLUTYRASPASPGLY-HISALASPMLEILETRPGLNLY	280
2947	DB	-----TCTGTATCTTGTGATCAAGATGAGTGTCCAAAGCCAGATGATTATAACAGT	3000
280	QY	SSERLEULEASNLYSARGTYRLEUGLNULEUTHYPRO-----	292
3001	DB	TGTTTACCTTTCGCCAACAGACCTCGCATTTTATCTCTCCCAAGAGAAACCGTAAGCGTT	3060
293	QY	-----HISLEULEUTHRGLUASPVASPVASPTYRASPASNPROGLYLEUSERCY	309
3061	DB	TCAGATTATACGGGGGTAAACCGTATCAAACCTAATCCAATGAATAATGAAAGCCAGTCT	3120
309	QY	SGLYPHEHISASPASPASPALAHISALAHISALAHISASNGLYLYSPROTRIPLEAS	329
3121	DB	TGGTTTTTAAGAGGGGGGTAT-CATTTTTCTGAACAACATTATATTCGGTGGTATTTTTGA	3179
329	QY	PLEUARGASNLYSARGTYRGLULEUARGALAGLUTRPLYSGLNPROPHEPROGLYPHEGL	349
3180	DB	ATTCAACACAAAAAATTGATATCCGT-----GATATGACATTTTCCCGCTTAT--	3228
349	QY	UALALEUARGVALHISLEUASNARGASNASPTYRHISHSASPGLULYSALAGLYASPA	369
3229	DB	-----TTAAGATCAACAGAAAACCGGATGATAG	3257
369	QY	AVAILGLUASNPHEPHEASNASNGLNTHRGLNASNALA-----ARGILEGLULEUARGHI	387
3258	DB	CAGTGGCTCTTTTATCCAAAGCAAGATTATGTTGCATATCAACGTATTCAG-----	3309
387	QY	SGLNPROILEGLYARG-----	393
3310	DB	-----GATGCGCCAGCGGTAACTATCGAAGTGGGCTTTATTTTCATGAACACCATAG	3362

Qy	393	uLysGlySerThrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThr	413
Db	3363	AAAACACGGCTAGGTATTTGAAATATATATACGAAATTAAGCAAAACGGGCATCATTTGA	3422
Qy	413	rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis--	428
Db	3423	CAAAAGCAGTGTAAAGTCTAATCAACAACAATCATATCTTGACAGTTATATGCAACATAC	3482
Qy	429	-----Ty	429
Db	3483	GCATTGCGTCTTTATCCTTAATCCAAGTAAGAATTGCGGCCCAACAGCTGATAAACCTTA	3542
Qy	429	rSerPheGlyValGlu-----	435
Db	3543	TTCAATCATATCTCTGTGTAGAATGTTTATAAGAAAAACATAATATGTTGCATTTGAA	3602
Qy	436	-----GlnAlaAsnTrp-----	439
Db	3603	TTTAGAGAAAAAATTCACAAAAATTCGCTTACTCATCAAAATTCCTCTCAATCTTGTTT	3662
Qy	440	-AspAsnPheThr-----LeuGluGlyValArgValG1	451
Db	3663	TGATGACTTTACTTCACGGCTTCAGCATAAAGATTATTAACTCGACGTGTACCGCTAC	3722
Qy	451	uLysGlnAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTrp	471
Db	3723	GGCAAGAGTATTCA-----GAGAAAGCTAATGAAACAAGAAATGTTTACAA	3773
Qy	471	sGln-----ProLeuProAspLeuGly-----	478
Db	3774	AAAAACAACCTTACTTATATACCCAAAACCAACAGTAGTTTGTAGTCAAGATCATTTGTA	3833
Qy	479	-----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs	492
Db	3834	TTATAAGGTTACTCTCTAATACAGAGACTGTAAAGTCGGTTAATTAAGGGAAAAA	3893
Qy	492	nTrpPheThrProGlnHisLysLeuSerLeu-----	503
Db	3894	TTATTATTTCGACGACGCAATATATGGCATTAGGAAATACGTTGATTTAGTTTAGG	3953
Qy	504	-----ThrAlaSerHisGlnGlu-----	509
Db	3954	TATTCGGTATGACGTATCTCGACAAAAGCTAATGAATCAACTATAGTGTGTGTAAT	4013
Qy	509	-----	509
Db	4014	TAAAAATTTCTCTTGGAACTACTGGTTATGTGCATAAAACCAACGGAATGCTTGATCTTC	4073
Qy	510	-----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys	522
Db	4074	TTATCGCCTTTCTACTCGATTAGAAATCTTAGTTTGTCTGAAATGTAT-----GGTTG	4127
Qy	522	sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAsSerAs	542
Db	4128	CGCGTATGTGGCAATTAATACGAGGTTTATGTAGTGAATTTAAGCCTGAAACATCTCG	4187
Qy	542	nAsnIleLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy	562
Db	4188	TAACCAAGAGTTTGGTCTCGCTCTAAAGGGGCAATTTGGTAAATTTGAGATCAGTCAATT	4247
Qy	562	rArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsn-----AspG1	578
Db	4248	TAGTAATGCTTATCGAAATCTTATCGCTTTGCTGAAGAACTTAATAAAATGGAACCTGG	4307
Qy	578	yArgGlyProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsn--G1	597
Db	4308	AAAGGCCAATTTATGGATATCATATATGCAAAAAATGCAAAATTAGTTGGCGTAATATAAC	4367
Qy	597	nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe-----	611
Db	4368	TGCGCAATTAGATTTTAATGGTTTATGGAACAGTATTCCTTACGCTGTGTTATGCAACATT	4427

```
QY 611 ----- 611
Db 4428 TGCTTATAACCGAGTAAAGTAAAGTCAAAAATCAATCTGGTTGGCCCTCGGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTGTGATGCAATTCAGCCAGC---CGTTATATCATTTGTTTAGCTATGA 4544
QY 624 P-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAAGTAATACTTTGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGluArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTCGTAGAAAACGT---GCATTGGGTAACAAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAAATCAACAAGAAAACCTTACTCGGCA-----TGGCATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGTATCGGGTTATTACATGCGGAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
Db 4723 -----AATATTATGCTTCGATTAGGGATATATAATTTATTCAA 4760
QY 713 nThrArgTyrGlyGluThrAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGCTATGTTACTTTGGGAAGCGGTGCGTCAAAACAGCACAGGTGCGGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATCTCGCTACGCGACATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 28
US-08-474-671-4
; Sequence 4, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
```

```
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160...2121, 2152...4890)
; US-08-474-671-4

Alignment Scores:
Pred. No.: 4,53e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 3 Gaps: 48

US-09-936-377-2 (1-758) x US-08-474-671-4 (1-5099)

QY 5 ThrLeuLysProLe-----ValLeuSerIleLeuLeuLeuAsnThrProLeu 20
Db 2155 ACTAAAAACCCCTATTTTCGCTAAAGTATTATTCTTCTTTTAAATTCATGCTATGTA 2214
QY 21 LeuAlaGlnAlaHis-----GluThrGlnSerValGly----- 32
Db 2215 AAAGCAGAAACTCAAGTATAAAGATACAAAGAAAGCTATATCATCTGAAGTGGACACT 2274
QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 2275 CAAAGTACAGAGATTCAGAAATTAGAACTATCTCAGTCTACTGCAGAAAAAATAAGAGAT 2334
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2335 CGTAAAGATAAATGAAGTA-----ACTGGACTTGGCAAAATTTATCAAAACTAGT 2382
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2383 GAAAGTATCAGCCGAGCAAGCAAGTATTAAATTCGTGATCTAAACACCTATGATCCAGGC 2442
QY 83 IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
Db 2443 ATTTCAGTTGTAGAACCAAGCCCGTGTGCAAGTCTCGATATTCTATTCTGGTATGGAC 2502
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 2503 AGAAATAGAGTT----- 2514
QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly 141
Db 2515 -----GCITTATTAGTAGTGTTTTACTCTCAACCGCAATCTTATGTAGTGCAGAAC 2565
QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 2566 CCITTATTGCTGCTTCAGGATATTCTGGCAGTGGTGCATTAATGAATTTGAATATGAA 2625
QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179
```

Db 2626 AATGTAAGGCGCGTGAA---ATAAGCAAGGGGGAGTCTCTGAGTATGTTAATGGA 2682
Qy 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198
Db 2683 GCATAGCTGTGTTCTGTACATATTCAAAGCAAAATCAGCAGCGGATATCTTTAGAGGAGAC 2742
Qy 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210
Db 2743 AAATCATGGGAATTCAAACTAAATGCTTATCAAGCAAAATAAAGGCTTTTACCACAT 2802
Qy 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
Db 2803 TCTTTAGCTGTAGCTGGAAAACAAGGGGATTTGACGGGTGCGCCATTTATATCTCAAGA 2862
Qy 222 AsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArg---Ala 240
Db 2863 AAT-----TCAATTGAAACCAAGTCCATAAAGATGCA 2895
Qy 241 ValLeuGlyTyrArgPheTyrArgThrTyrSerAspArgAspGlnTyr 260
Db 2896 TTAAGGCGGTACAAAGT-----TATCATCGATTAAATCGCCAAACCGAGGATCAA--- 2946
Qy 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTyrGlnLys 280
Db 2947 -----TCTGCATATTTGTGATGCAAGATGAGTGTCCAAAGCCAGATGATATACAGT 3000
Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
Db 3001 TGTTTACCTTTTCGCAAAACGCTCGGATTTTATCTCCCAAGAGAAACCGTAAGCGTT 3060
Qy 293 -----HisLeuLeuThrGluLysValAspTyrAspAsnProGlyLeuSerCy 309
Db 3061 TCAGATTATACGGGGCTTAACCGTATCAACCTTAATCCAAATGAATATGAAGCCAGTCT 3120
Qy 309 sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAs 329
Db 3121 TGGTTTTTAAGAGGAGGTAT-CATTTTCTGCAACAACATTATATATGTTGGTATTTTGA 3179
Qy 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheGlyPheG 349
Db 3180 ATTCAACAACAAAAATTTGATATCCGT-----GATATGACATTTCCCGCTTAT-- 3228
Qy 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlyLysAlaGlyAspAl 369
Db 3229 -----TTAAGATCAACAGAAAACGGGATGATAG 3257
Qy 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
Db 3258 CAGTGGCTCTTTTATCCAAAGCAAGATTATGTTGCATATCAACGTATTGAG----- 3309
Qy 387 sGlnProIleGlyArg----- 393
Db 3310 -----GATGGCGAGCGGTTAACTATGCAAGTGGGCTTTATTTTCGATGAACACCATAG 3362
Qy 393 uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
Db 3363 AAAACAGCGTGTAGTATTGAATATATTTACGAAATAAGAACAAAGCGGCATCATGA 3422
Qy 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
Db 3423 CAAAGCAGTGTAAAGTGCTAATCAACAAACATCATCTTGCACAGTTATATGCAACATAC 3482
Qy 429 ----- 429
Db 3483 GCATTGCGAGTCTTATCTCTAATCCAGTAAGATTGCGGCCCAACACGATGAACACCTTA 3542
Qy 429 rSerPhePheGlyValGlu----- 435
Db 3543 TTCATCTACTCATCTGTAGATAAATGTTTATAAGAAAAACATAATATGTTGCAATTGAA 3602
Qy 436 -----GlnAlaAsnTrp----- 439

Db 3603 TTTAGAGAAAAAATTCACAAAAATTTGGCTTACTCATCAAAATTTCTTCAATCTTGGTTT 3662
Qy 440 -AspAsnPheThr-----LeuGluGlyGlyValargValGI 451
Db 3663 TGATGACTTTTACTTTCAGCGCTTCAGCATAAAGATTATTAACTCGACGTTTACCCTAC 3722
Qy 451 uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys 471
Db 3723 GCGAAAGAGATTATTCA-----GAGAAAGCTAATGAAACAAGAAAGATTTGGTTACAA 3773
Qy 471 sGln-----ProLeuProAspLeuGly----- 478
Db 3774 AAAACAACCTTACTTATACCCAAACCAACAGTAGCTTTTGTAGTACAGATCATTTGTA 3833
Qy 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGTAACTCTCTTAATTACAGAGACTGTAAAGTGGCTTAAATTAAGGGGAAAAA 3893
Qy 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu----- 503
Db 3894 TTATTATTTCGACGACGCAATAATATGTCATTAGGAAATACGTTGATTAGGTTTAGG 3953
Qy 504 -----ThrAlaSerHisGlnGlu----- 509
Db 3954 TATTCGGTATGACGTATCTCGCACAAAAGCTAATGAATCAACTATTAGTGTGGTAAATT 4013
Qy 509 ----- 509
Db 4014 TAAAAATTTCTTCTTGGAACTATGTTGTCATAAAACCAACGGAATGCTTGAATCTTC 4073
Qy 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys 522
Db 4074 TTATCGCCTTTCTACTCGAATTAGAAATCTTAGTTTTCGTGAATGTAT-----GGTTG 4127
Qy 522 sHisValAlaThrAsnThrPheGluValGlyLysAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 CGCGTATGCTGGCAATAATAGCGAGTTTGTAGTAAATTTAAGCCCTGAAACATCTCG 4187
Qy 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyr 562
Db 4188 TAACCAAGAGTTTGGTCTCGCTCTAAAGGGGATTTTGGTAATATAGATCAGTCATTT 4247
Qy 562 rArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsn-----AspGI 578
Db 4248 TAGTAATCTTATCGAAATCTTATCGCCTTTTGTGCAAGAACTTAATAAAATGGAACCTG 4307
Qy 578 YArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn---GI 597
Db 4308 AAAGGCCAATTATGGATATCATATGCAAAAATGCAAAAATTAGTTGGGTAAATATAAC 4367
Qy 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db 4368 TCGCAATTAGATTATTAATGTTTATGGAACGTTATCCCTACGGTTGTTATGCAACATT 4427
Qy 611 ----- 611
Db 4428 TGCTTATAACCGAGTAAAGTTAAAGATCAAAAAATCAATGCTGTTGGCTCCGTAAG 4487
Qy 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTATGATGCCATTTCAGCCCGCAGC-----CGTTATATCATTTGTTAGGCTATGA 4544
Qy 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAAAGTAATACTTGGGGAATTAATACAACTGTTTACTCAATCAAAAGCAAAATCTCA 4604
Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheLeIleAlaGlnAspAs 653
Db 4605 AAATGAATTTCTAGGAAAAACGT---GCATTTGGGTAAACAAT-----TCAAGGATGT 4652
Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAAGAAACTTACTCGGGCA-----TGGCATATC----- 4689


```

RESULT 30
US-08-897-438-4
; Sequence 4, Application US/08897438
; Patent No. 6262016
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,438
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160..2121, 2152..4890)
US-08-897-438-4
Alignment Scores:
Pred. No.: 4,538-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 3 Gaps: 48
US-09-936-377-2 (1-758) x US-08-897-438-4 (1-5099)
QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuLeuAsnThrProLeu 20

```

Db 3121 TGGTTTAAAGAGGGGAT-CAITTTTGAACAACATTATATGCTGGTATTTTGA 3179
 Qy 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheCl 349
 Db 3180 ATTACACACACAAAATTTGATATCGT-----GATATGACATTTTCCCGCTTAT-- 3228
 Qy 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369
 Db 3229 -----TTAGATCAACACAGAAAACCGGATGATAG 3257
 Qy 369 aValGluAsnPheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHis 387
 Db 3258 CAGTGGCTCTTTTCCAAAGCAACGATATTGCTGATATCAACGTTATTCAG----- 3309
 Qy 387 sGlnProIleGlyArg-----Le 393
 Db 3310 -----GATGGCCGAGCGGTTAACTATGCAAGTGGGCTTTATTTTCGATGAACACCATAG 3362
 Qy 393 uLysGlySerTrpGlyValGlnTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
 Db 3363 AAAACAGCGGTAGGTATTGAATATATTACGAAAATAAGAACAAAGCGGCATCATTTGA 3422
 Qy 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
 Db 3423 CAAAGCAGTGTAAAGTCTTAATCAACAACATCATACTTTGACAGTTATATGCAACATAC 3482
 Qy 429 -----Ty 429
 Db 3483 GCATTGCGCTTTTATCTTAATCCAAAGTAAGATTGCGCCCAACACGCTGATAAACCTTA 3542
 Qy 429 rSerPhePheGlyValGlu----- 435
 Db 3543 TTCATACTACTTCTGATAGAAATGTTTATAAGAAAAACATAATATGTTGCAATTTGAA 3602
 Qy 436 -----GlnAlaAsnTrp----- 439
 Db 3603 TTTAGACAAAATAATCAACAATTTGGCTTACTCATCAATTTGCTTCAATCTTGTTT 3662
 Qy 440 -AspAsnPheThr-----LeuGluGlyValArgValG 451
 Db 3663 TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTTAACTCGAGCTGTTACCGCTAC 3722
 Qy 451 uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTrpLys 471
 Db 3723 GGCNAAGAGTATTTCA-----GAGAAAGCTTAATGAACAAGAAATGGTTACAA 3773
 Qy 471 sGln-----ProLeuProAspLeuGly----- 478
 Db 3774 AAACAACCTTACTTATACCCMAAACCAACAGTAGGTTTGTAGTACAAAGATCATTTGTA 3833
 Qy 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
 Db 3834 TTATAAAGGTAACTCTCTAATTACAGAGACTGTAAAGTCGCGTTAATTAAAGGGAAAAA 3893
 Qy 492 nTriPyrPheThrProGlnHisLysLeuSerLeu----- 503
 Db 3894 TTATTATTTCGACGACGCAATATATGCGATTAGGGAAATACGTTGATTAGTTTAGG 3953
 Qy 504 -----ThrAlaSerHisGlnGlu----- 509
 Db 3954 TATTGCGTATGACGTATCTCGCACAAAGCAATGAATCAACTATTAGTTGTTGTAAT 4013
 Qy 509 ----- 509
 Db 4014 TAAAAATTCCTTGGATACTGTTATGTCATAAAAACCAACGAATGGCTTATCTTTC 4073
 Qy 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys 522
 Db 4074 TTATCGCTTTCTACTCGAATTTAGAAATCTTAGTTTTCGCTGAAATGTAT-----GTTG 4127
 Qy 522 sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
 Db 4128 GCGGTATGGTGGCAATAATAGCGAGGGTTTATGTAGGTAAATTTAAAGCCGTAACATCTCG 4187

Qy 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTrpAsnLeuAlaLeuTy 562
 Db 4188 TAAACAGAGTTTGGTCTCGCTCTAAAGGGGATTTTGGTAATATTAGATCACTCATTT 4247
 Qy 562 rArgAsnArgPheGlyAsnTrpIle-----TyrAlaGlnThrLeuAsn-----AspG 578
 Db 4248 TAGTAATGCTTATCGAAATCTTATCGGCTTTTGCCTGAAGAACTTAATAAAATGAACCTGG 4307
 Qy 578 yArgGlyProLysSerIleGluAspAspSerGluMetIleLeuValArgTyrAsn---G 597
 Db 4308 AAAGCCCAATTATGATATCATATCAATAATGCAAAATGCAAAATAGTTGCGGTAATAAATAC 4367
 Qy 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
 Db 4368 TGGCAATTAGATTTAATGCTTTATGMAACGTAATTCCTACGCTGGTGGTATGCAACATT 4427
 Qy 611 ----- 611
 Db 4428 TGCTTATAACGAGTAAAGATTAAAGATCAAAAATCAATGCTGTTGGCTCCGTAAG 4487
 Qy 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
 Db 4488 CAGTTATTTATTTGATGCTCCATTCAGCCGAGC---CGTTATATCATTTGCTTTAGCTATGA 4544
 Qy 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
 Db 4545 TCATCCAAAGTAATACTTGGGGAATTAAATACATGTTTACTCAATCAAAAGCAAAATCTCA 4604
 Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
 Db 4605 AAATGAATTTGCTAGAAACGTT---GCATTGGGTAACAAAT-----TCAAGGGATGT 4652
 Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
 Db 4653 AAAATCAACAAGAAACTTACTCGGCA-----TGGCATATC----- 4689
 Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
 Db 4690 -----TTAGATGTATCGGGTTATTATACATGGCGAATAAA----- 4722
 Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
 Db 4723 -----AATATTATGCTTCGATTAGGATATATAATTTATTTCAA 4760
 Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
 Db 4761 CTATCGCTATGTTACTTGGGAAGCGGTGCTCAACACAGCAACAGGTGGGTCAATCAACA 4820
 Qy 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
 Db 4821 TCAAAATGTTGTAGCTATATCTCGCTACGAGCATCA-----GGACGAAACTA 4868
 Qy 750 eThrGlyGlyValAsnValLysPhe 758
 Db 4869 TACCTTAACATTAGAAATGAATTC 4893
 RESULT 31
 US-08-637-654-4
 ; Sequence 4, Application US/08637654
 ; Patent No. 6358727
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M
 ; APPLICANT: Harkness, Robin E
 ; APPLICANT: Schryvers, Anthony B
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Mardin, Andrew D
 ; APPLICANT: Klein, Michel H
 ; TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney
 STREET: 6th Floor, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/637,654
 FILING DATE: 05-AUG-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/CA94/00616
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-595
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5099 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join(160..2121, 2152..4890)
 US-08-637-654-4

Alignment Scores:
 Pred. No.: 5099
 Score: 180.50
 Percent Similarity: 33.30%
 Best Local Similarity: 19.52%
 Query Match: 4.47%
 DB: 4

US-09-936-377-2 (1-758) x US-08-637-654-4 (1-5099)

QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
 Db 2155 ACTAAAGAAACCTATTTTCGCTAAGTATATTCTTCTTTTAAATTCATGCTATGTA 2214
 QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
 Db 2215 AAAGCAGAACTCAAAGTATAAAGATACAAAGAAAGCTATATCATCTGGAAGTGGACACT 2274
 QY 33 -----LeuGluThrValThrValValGly-----LysSerArgPro 44
 Db 2275 CAAAGTACAGAGATTCAGAACTTACAACTATCTCAGTCACTGCAGAAATAAAGAGAT 2334
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
 Db 2335 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATTTATCAAACTAGT 2382
 QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
 Db 2383 GAAAGTATACCGCGAAGACAGTATTAATATTCGTGATCTTAACACGCTATGATCCAGGC 2442
 QY 83 IleHisAlaSerGlnThrGlyGlyAlaSerAlaProVal-----IleArgGlyGlnThr 101
 Db 2443 ATTTTCAGTTGTGATGAAACAAAGCGCGTGCAGAGTTCTCGGATATTTCTGTTGATGAC 2502
 QY 102 GlyArgGlnIleLysValLeuAsnHisGlyGlyGluThrGlyAspMetAlaAspPheSer 121
 Db 2503 AGAATAGAGTT----- 2514

QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGly 141
 Db 2515 -----GCTTTATTAGTAGATGTTTACTCTAAACGCAATCTTATGATGTCAAAGC 2565
 QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
 Db 2566 CCTTTAGTTGCTGCTCAGGATATTTCTGGCACTGTGCAATTAATGAATGAATGAA 2625
 QY 160 AspGlyLysIleProGluLysMetProGluAsnGlnValSerGlyGluLeuGlyLeuArg 179
 Db 2626 AATGTAAAGCGCGTCGAA-----ATAAGCAAGGGGGGAGTTCTTCTGAGTATGGTAAATGGA 2682
 QY 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly 198
 Db 2683 GCCTAGCTGTTCTGTACATTTCAAAGCAAAATCAGCAGCGGATATCTTAGAAGGAGAC 2742
 QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210
 Db 2743 AAATCATGGGAATTCAAAATAAAATGCTTATTCTACGACAAAATAAAGGCTTTACCCAT 2802
 QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
 Db 2803 TCTTTAGCTGTAGCTGGAATAAAGGGGATTTGACGGGGTGGCCATTTATCTCAACGA 2862
 QY 222 AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala 240
 Db 2863 AAT-----TCAATTTGAAACCCCAAGTCCATAAAGATGCA 2895
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgThrTyrSerAspArgArgAspGlnTyr 260
 Db 2896 TTAAGAGCGGTACAAAGT-----TATCATCGTAAATATCGCCAAACCGAGAGATCAA 2946
 QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTyrGlnly 280
 Db 2947 -----TCTGCATACTTTGTGATGCAAGTGTGCCAAAGCAGATGATTATACAGT 3000
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
 Db 3001 TGTTTTACCTTTTCGCCAAACGACCTGCGATTTTATCTCTCCAAAGAGAAACCGTAAGCGTT 3060
 QY 293 -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys 309
 Db 3061 TCAGATTATACGGGGCTAACCGTATCAACCTATCCATGATGAATATGAAGACGAGTCT 3120
 QY 309 sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAs 329
 Db 3121 TGGTTTTTAAGAGGAGGTAT-CATTTTCTGCAACAACATTTATTTGGTGGTATTTTGA 3179
 QY 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349
 Db 3180 ATTCACACAAACAAATTTGATATCCGT-----GATATGACATTTCCCGCTTAT 3228
 QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369
 Db 3229 -----TTAAGATCAACAGAAATAAAGCGGATGATAG 3257
 QY 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgH 387
 Db 3258 CAGTGCCTCTTTTATCCAAAGCAAGATTTGCTGCATATCAACCGTATTGAG----- 3309
 QY 387 sGlnProIleGlyArg----- 393
 Db 3310 -----GATGCCGAGCGTTAACTATCAAGTGGCTTTTATTTTCGATGAACACCATAG 3362
 QY 393 uLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
 Db 3363 AAAACAGCGGTAGGTATTGAATATATTACGAAATAAAGAAACGCGGATCATTTGA 3422
 QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
 Db 3423 CAAGCAGTGTAAAGTGCTAAATCAACAAACATCATCTTACAGTTATATGCAACATAC 3482

QY 429 -----TY 429
Db 3483 GCATTGAGCTTTATCTATCCAGTAAGTAAGTAAGTGGCCCAACACAGCTGATAAACCTTA 3542
QY 429 rSerPheGlyValGlu 435
Db 3543 TTTATCTATCTCTGATGAGTAATGTTTATAAGAAAAACATAATATGTTGCAATGAA 3602
QY 436 -----GlnAlaAsnTrp 439
Db 3603 TTTAGAGAAAAAATTTCAACAAATTTGGCTTACTCATCAAAATCTCTCAATCTTGTTTT 3662
QY 440 -AspAsnPheThr 451
Db 3663 TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTAACTCGAGCTGTACCGCTAC 3722
QY 451 uLysGlnLysAlaSerIleAArgTyrAspLysAlaLeuIleAspArgGluSerTyrTyr 471
Db 3723 GGCAGAGAGATTTCATCA-----GAGAAAGCTTAATGAACAAAGAAATGCTTACAA 3773
QY 471 sGln-----ProLeuProAspLeuGly 478
Db 3774 AABACACCTTACTTATATCCCAACCAACAGTAGGTTTGTAGTACAGATCATTCTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGGTAACTCTCTCTATTAATACAGAGACTGTAAGTGCAGTTAAATTAAGGGAATA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db 3894 TTATTATTTCGACGACGCAATATATGCTTACGTAATGAGGAAATACGTTGATTAGGTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu 509
Db 3954 TATTCGGTATGACGTATCTGCACAAAGCTAATGAATCAACTATTAGTGTGTAAT 4013
QY 509 ----- 509
Db 4014 TAAATAATTTCTTGGAACTAGTGGTATTGTTCATAAACCAACGAAGTGGTGTCTTCTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGly 522
Db 4074 TTATCGCTTTCTACTGGATTTAGAAATCTAGTTTCTGTAATGTAT-----GGTTG 4127
QY 522 shiValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGGTGCATAATAGCGAGGTTTATGTAGTAAATTAAGCTCGAAACATCTCG 4187
QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAAACAAGAGTTTGTCTCGCTCTAAAGAGGAGTTTGTGTAATATTGAGATCAGTCAATT 4247
QY 562 rArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsn-----AspG 578
Db 4248 TAGTAATGCTTATCGAAATCTTATCGCTTGTCTGAGAACTTAATAAAATGGAACCTGG 4307
QY 578 yArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn---G 597
Db 4308 AAGGCCAATTATGATATCATATGACCAAAATGCAAAATAGTTGGCGTAATATAAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe 611
Db 4368 TCGCAATTATGATTTTAAATGTTTATGGAACAGCTATTCCTACGTTGGTATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCCTTATACCGAGTAAAGTTAAAGATCAAAAAATCAATGCTGTTGGCTCGCTGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTTTATGATGCGCAATTCAGCCAGC---CGTTATATCATGTTGGTTAGGCTAGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633

Db 4545 TCATCAAGTAATACTTGGGAATTAATACATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTTGTTAGGAACAGT---GCATTGGGTAAACAAT-----TCAAGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAAAGAACTTACTCGGCA-----TGGCATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGTATCGGGTTATTATACATGCGCAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 4723 -----AATATTATGTTTCGATTAGGATATATAATTTATTCA 4760
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys-----AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGCTATGTACTTGGGAACGGTGCCTCAACAGCAGCAAGGTGCGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATCTACTCGCTACCGCAGCATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACTTAGAATGAATTC 4893

RESULT 32
US-08-649-518-4
; Sequence 4, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Van-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,518
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968

Db 4723 -----AATATTATGCTTCGATTAGGATATATAATTATTATCA 4760
 Qy 713 nThrArgTyrGlyGluThrAsnTrpTyrVallys-----AlaAspAsnLeuLeuAsn----- 730
 Db 4761 CTATCGCTATGTTACTTTGGAGCGGTGCGTCAACAGCACCAAGGTGCGGTCAATCAACA 4820
 Qy 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
 Db 4821 TCAAAATGTTGGTAGCTATCTACTCGCTACGACGATCA-----GGACGAACACTA 4868
 Qy 750 eThrGlyGlyValAsnVallysPhe 758
 Db 4869 TACCTTAACATTAGAAATGAAATTC 4893
 RESULT 33
 US-09-252-991A-14238
 ; Sequence 14238, Application US/09252991A
 ; Patent NO. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 14238
 ; LENGTH: 2271
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-14238
 Alignment Scores:
 Pred. No.: 1.36e-08 Length: 2271
 Score: 180.00 Matches: 179
 Percent Similarity: 33.22% Conservative: 106
 Best Local Similarity: 20.86% Mismatches: 276
 Query Match: 4.46% Indels: 297
 DB: 4 Gaps: 44
 US-09-936-377-2 (1-758) x US-09-252-991A-14238 (1-2271)
 Qy 8 ProfileValLeuSerIleLeuLeu-----IleAsnThrProLeuLeuAlaGlnAla 24
 Db 73 CCCCTGGCGTTCCTCCGTCCTCTCGCTTCGTTTCGCTCTCGCTCGCTGCGT---GCC 129
 Qy 25 HisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
 Db 130 GATCCGCTGGACGACGAG-----ATGGTGGTGATCGGCTCGCGCGCGCG 174
 Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspIleIleSerGlyAsp 64
 Db 175 ACAGGATACGACGATGGTCCCGCACCGCTCGGTGGATGACGACGC---GAGCAACTGGAC 231
 Qy 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly---ValProGlyIle 83
 Db 232 CAGCAGACCCAGCGCGCGTGCCTGAAGAGGCGTGGGACAAATGATCCCGGGCGCTG 291
 Qy 84 HisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArg 103
 Db 292 GATATCGGCTCCCGAGGTCGTACCAACAGCGCACCACTGGCG-----GGACGC 342
 Qy 104 ArgIleLysValLeu-----AsnHisGlyGluThrGlyAspMetAlaAsp 119
 Db 343 AGCGTACTGTTGATGATCGACGGGGTCTCCTGACAGTTTCACGGGGCATCACCCTGGCAG 402
 Qy 120 PheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu 139
 Db 403 TTCGAC-----TCCATCGACCCCATCAACATCGAGCGCATCGAGGTGATG 447


```

US-09-252-991A-5404
/ Sequence 5404, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 5404
/ LENGTH: 2403
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5404

Alignment Scores:
Pred. No.: 1,498-08 Length: 2403
Score: 180.00 Matches: 173
Percent Similarity: 37.02% Conservative: 103
Best Local Similarity: 23.45% Mismatches: 264
Query Match: 4.46% Indels: 218
DB: 4 Gaps: 43

US-09-936-377-2 (1-758) x US-09-252-991A-5404 (1-2403)
Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 AAATCACCAGCGCTCGCG-ACCAGCAGCGACCTGGTCTCCGCCACGCGCCAGTCCACCGTC 492
Qy 61 IleSerGly-----AspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu 77
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
493 ATCAGCATGCCCACTGAGGAGCTCGCCAGGGTCCGACAGCTCGCCACCGTGGT 552
Qy 78 ---AspGlyValProGlyIleHisAlaSer-----GlnTyrglyGlyGly 91
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
553 GCCAAGCGGTGCGCGGATGTCGATTCAGCGCGCACCACCATCACCGAATAC----- 603
Qy 92 AlaSerAlaProValIleArgGlyGlnThr-----GlyArgGlyLeuValLeuAsn 109
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 -----GGCCAGACCTCGCGCGGCGCAGCATGCTGGTGGTGGT 642
Qy 110 HisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp 129
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
643 GACGGCGTGGCTGAAACACCAACCGGACTCTCTCGCGCAACCTGGCC---AACATCGAC 699
Qy 130 ThrAlaLeuSerGlnValGluLeuLeuArgGlyProValThrLeuLeuTyrsSer 149
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
700 CCGCGCTGATCGAGCGGATGAGGTATCTCGTGGCAGC---AGCGCCATCTACCGCAGC 756
Qy 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu 169
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
757 GGCGCCCGCGCGGATCATCTCATCACC-----ACCGTCCGCGCC 798
Qy 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn---LeuGluLysLeu 188
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
799 GGCGGC---GAGAACCGCGGAAACCCGCTCAGCGCCACCTCGCGCTGACCGCGCTG 855
Qy 189 ThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu 208
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
856 GGCAGCAT-----GGCTCGCGCGCAGTTTC----- 882
Qy 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAsp 228
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
883 -----CAGCAATA-CTTCGCGCGCTCCCTGGG 908
Qy 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 248
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 909 GGCGCTGACTATTCTGACTTCGGCACCGCTCA-----CGTCGCGCTTC 956
Qy 249 TyrArgArgThrTyrsSerAspArgArg-----AspGlnTyrglyLeuProAlaHis-Se 266
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
957 CTACGAGCGCCATGGCAGCGCATCGCCCGGAAACCCAGCAGCGCGACCTGTCGACTC 1016
Qy 266 rHisGlnTyrgly-----AspAspCysHisAlaAspI 276
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1017 GAAGCTCTACAACATCGCGGCAAGCTCGGCTCGCATCGACGAG----- 1062
Qy 276 eIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrsProHisLeuLeuTh 296
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1063 -----AACCGCGCTCGAGCTCGCCCTCAGCAGCTACGACG 1100
Qy 296 rGluGluAspValAspTyrAspAsn-----ProGlyLeuSe 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1101 CGCGCAGGACACCGACTACGCCACCGCAGCGGCTGCCAGGCTGCCGCGGC----- 1155
Qy 308 rCysGlyPheHisAspAspAlaHisAlaHisAsnGlyLysProTrpI 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1156 -----TCGTCCTCCGCAACCGCATCAAGGCTCGGAG---CT 1190
Qy 328 eAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPh 348
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1191 GGACGAGCGAAACCGCATCCCAACACCTCGCGCAACCTCGAG-----TACGAGAAGCT 1244
Qy 348 eGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAsp----- 363
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1245 CGACATCTCGCGCAGCGGCTCTCGCGCAGCTCTACTACCGTGACTATTTACCGCGTT 1304
Qy 364 -----GluLysAlaGlyAspAlaValGluAsnPhePh 374
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1305 CACTCCGTTCCAGCGCGCGCTCTCCACCGCGGCGCAACCTCGCAGCATGTCGA 1364
Qy 374 eAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLy 394
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1365 GAACAGCGAAGTGTTCGCGCAGCGCTGACCTCGCGC---ACGCGCTGGCGGAAGCGG 1421
Qy 394 sGlySer-----TrpGlyValGlnTyrglyGlnLysSer----- 406
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1422 CAATACCGAAGTGTGTTCGCGCGCGACTACACCGAGGACGCGCAGCATCGCGCTCGA 1481
Qy 407 -----SerAlaLeuSerAlaThrSerGluAlaVal-----Ly 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 CGTGTTCAGCGCGCGCTAGCAGCGCGCGGCTGTCTTCGACAGATCGGCA 1541
Qy 417 sGlnProMetLeuLeuAspAsnLysValGlnHisTyrsSerPheGlyValGluGlnAl 437
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 GCTCACCTACATGCGCGCTAGCAGCGCGCGCGCGCTCGCGCGCTTCGCGCGTTCGACGA 1601
Qy 437 aAsnTrpAsp---AsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSe 456
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1602 TCGCTTCGACGAACTGTGTGATCGAGCGCGCTCGCTACGAATACTCCACGCGCGA 1661
Qy 456 rIleArgTyrsAspLysAlaLeuIle-----AspArgGluAs 468
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1662 -ATTCGAGACTTCGCTCGCTCTCCGAGTCCAGGCGCGCTCGCGCTGACCGTCAAGG 1720
Qy 468 nTyrtTyrsGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAl 488
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1721 G-----CGCGGACTTCGACTACGAGCGCGTGTTCGAACCTCGGCATCTGCTATTTCG 1774
Qy 488 aLeuSerGlyAsnTrpTyrsPheThrPheGlnHisLysLeuSerLeuThrAlaSerHisG 508
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1775 C-----GGTGGCTGG-----CCAGGAATCTATGCTCTCTT-----CA 1807
Qy 508 nGluArgLeuProSerThrGlnGluLeuTyrsAlaHisGlyLysHisValAlaThr-Asn 528
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1808 GCCAGGGCTTCAGCT-----GCCGATGTGGCATCCAGCTCGGCAACG 1852
Qy 528 hr-----PheGluValGlyAsnLysHisLeuAsnLysGluArgSerHisnAsnIleG 545
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1853 CCCGCGGTGGCTTCGATATCGGCTCTCTCGAACCTCGAACCGGTGAAGACCACTACG 1912

```

```
QY 545 lufeuAla---LeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyr----- 562
Db 1913 AACTCGGCTGGCGGCGCCATCGGCGGCAACACCTGGCGACCCCTGGCGCTGTTCATCA 1972
QY 563 --ArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProL 582
Db 1973 CCACCTCAAGCTGGGCGAC-----GTGAGAGCTTCACACACGGC----- 2013
QY 582 ysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspP 602
Db 2014 -----CTGATCTCTCACCGCCACCAAGGAGCGC-----A 2041
QY 602 heTyrGlyAlaGluGlyGluIle---TyrPheLysProThrProArgTyrArgIleGlyV 621
Db 2042 TCTACGGCGTGGAGCGAGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2101
QY 621 alSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluA 641
Db 2102 GCAGCGCCACCTGGATGCTGCGGCGCGAGGAG-----CCGAGCGGCAAGG 2146
QY 641 spAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProA 661
Db 2147 AC-----TGGCAGGACATGACCGGCTACCGGCTGCGCGC 2179
QY 661 laAlaArgLeuGlyPheHis-----LeuL 669
Db 2180 CGCTGAAGCTGACCGCCTACCTGCAATACAAAGCGGATGCGACTGGAAACACCGCCTGC 2239
QY 669 ysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnA 689
Db 2240 AGGCACCTCTTCGACTCCAGGACTACCTGCTCGAGCGGCGTGGAAAGCTTCGGCGCGC 2299
QY 689 snLysLeuAlaArgTyrGluThrArgThrProGly-----HisHis 702
Db 2300 GCGAGTGCAGCAGCTACACACCGGT--CGACCTGCTGAGCGAGTACCGCATCAC 2351

RESULT 35
US-09-307-973A-2
; Sequence 2, Application US/09307973A
; Patent No. 6432686
; GENERAL INFORMATION:
; APPLICANT: BULTHIUS, BEN
; APPLICANT: GATENBY, ANTHONY A.
; APPLICANT: TRIMBUR, DONALD E.
; APPLICANT: WHITED, GREGORY
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
; TITLE OF INVENTION: 1,3-PROPANEDIOL BY RECOMBINANT
; TITLE OF INVENTION: ORGANISMS COMPRISING GENES FOR
; TITLE OF INVENTION: VITAMIN B12 TRANSPORT
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; ADDRESSSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 925 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/307,973A
; FILING DATE:
; CLASSIFICATION: 435
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,190
; FILING DATE: JUNE 30, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1245
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1844 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-307-973A-2

Alignment Scores:
Pred. No.: 2,52e-08 Length: 1844
Score: 176.00 Matches: 139
Percent Similarity: 34.31% Conservative: 83
Best Local Similarity: 21.48% Mismatches: 240
Query Match: 4.36% Indels: 186
DB: Gaps: 33

US-09-936-377-2 (1-758) x US-09-307-973A-2 (1-1844)
QY 29 GlnSerValGlyLeuGluThrValThrValGlyLysSerArgProArgAlaThrSer 48
Db 61 CAGGACACTAGCCCGGATACCTGCTGCTCCACCGCAACCGCTTTTCAGCAGCGCGCAGC 120
QY 49 GlyLeuLeuHisThrSerThrAlaSerAspLysLlelleSerGlyAspThrLeuArgGln 68
Db 121 GCGGTTCTGGCGCCCGTTCACCATC-----GTGACGCGTCAGGATATTGAACGCTGG 171
QY 69 LysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyr 88
Db 172 CAATGACCTCGTAAATGATGTTCTGCGCGCTTTGCTGCGCTGATATTGGCAGAGC 231
QY 89 GlyGlyGlyAla-----SerAlaProValIleArgGlyGlnThrGlyArgArgIleLys 106
Db 232 GCGCGCGCGCAGCAAAACTCTCTCATTTTCATTCGCGCAACCAACTCCAGCCATGACTG 291
QY 107 Val-----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 292 GTATTGATTGACGCGCTGCTGAATTTAGCAGCGCTGAGCGGTCC--GCCGATCTC 348
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLeuArg 140
Db 349 AGC-----CAGTTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 394 GGTTCGCGCTCGCTATTATTATGTTCCGATGCTATCGCGCGCGCTAGTGAATATC----- 447
QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 180
Db 448 -----ATTACGACGCGCAT 462
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGly-----GlyIleAsnIleGlyLeuGlyLys 199
Db 463 AACCCAGGCACA-----GAATTAACCGCTGATGGGGAAGCAAT-----AGTACCAGC 510
QY 200 AsnPheValLeuHisThrGluGlyLeu----- 208
Db 511 AATTACGACATCTCGACGCAACAGCAACTTGGCGAAATCACGCGCGCGCGCTGATCGGC 570
QY 209 -----TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225
```

```

Db 571 GATTACCAATACACCAAGGGTTTACGGTGGTAGCGAAAGGCGGTACCGGATCGAGCG 630
QY 226 LeuProAspSerProArgArgPheAlaAenGlyGlnHisArgAlaValLeuGlyTrpArg 245
Db 631 CAGCTTGACGGGACGGCTTTTGTAGTAAACGCCTTTATGGCGCTTA----- 678
QY 246 LysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHis 265
Db 679 -----GAGCATACCTTTCTGATCGCTGGAGCGGATTC-----GTGCGT 717
QY 266 SerHisGlnTyrAspCysHisAlaAspIleIleTrpGlnLysSerLeuLeuAsnLys 285
Db 718 GGTATGGCTACGATAAC----- 735
QY 286 ArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnPro 305
Db 736 -----CGTACCGATTACGAC----- 750
QY 306 GlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAsnGlyLys 325
Db 751 -----GCCTATTACTCGCGGCTCG 771
QY 326 ProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPhe 345
Db 772 CCGCTGATTATACACGC-----AAACTTTATAGCCAAAGCTGGAGCGCGGCTGCACATT 828
QY 346 ProGlyPheGluAlaLeuArgValHisLeu-----AsnArgAsnAspTyr 360
Db 829 AATGCG-----GAAAGTATTCACTCAGCTCGTGTTCAGCTATAGCCACAGATAAGATTAC 885
QY 361 HisHisAspGlnLysAlaGly-----AspAlaValGluAsnPhePheAsnGlnThr 378
Db 886 AACTATGATCCGCACTATGCGCGGTATGATACCTCCGCCACGCTGATGA-GATGAACA 944
QY 379 GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGly 398
Db 945 GTACAAATGTTCAATGGAC-----CAACAGTGTGTCTGTTGGGACGGTAATGTTGGG 995
QY 399 ValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGln 418
Db 996 GCGGGCGGTAGACTGGCAGAAACAGACTACACGCCAGAGTACCGGCTATGTG----- 1046
QY 419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
Db 1047 CCGGAGGATATGACACAGCTAATACCGGGTTTACCTGACAGATTACACAGTTGGT 1106
QY 439 TrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArg 458
Db 1107 -----GACTTCATCTGTGAAGCGCGCGCGCGAGTACGACAACTCC----- 1148
QY 459 TyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGly 478
Db 1149 -----CAGTTTGGT 1157
QY 479 AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGln 498
Db 1158 CGTCAT-----GGTACATGGCAACACCGCGCGGATGGGAGTTTATAGAAGT 1205
QY 499 HisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyr 518
Db 1206 TATCGCTTTATGCTCTCTACGAGAACTCTACAAAGCGCTAATTTGGCCAACTGTAT 1265
QY 519 AlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLys 538
Db 1266 GGTAT-----TACGGTAAATCCGAACCTGAATCCT 1295
QY 539 GluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu-----GlyAspArg 554
Db 1296 GAAAGAGTAAACAGTGGAGGGCA-----TTTGAAGGCTAACCGCTGCGCTCAGC 1349
QY 555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThr 574

```

```

Db 1350 TGCGGT-----ATTTCAGTTATCGTAAC-----GAT 1376
QY 575 LeuAsnAspGlyArgGlyProLysSerIleGluAspSerGluMetLysLeuValArg 594
Db 1377 ATTAATGAC-----ATGATCGATTATGACGATCATCTGCAGAAATAT--- 1418
QY 595 TyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThr 614
Db 1419 TACACGAGAGTNAAGCGCGCATTAAGGTATTAGGCGACCGCGAATTTTCATACCGGA 1478
QY 615 ProArgTyrArgIleGlyValSerGlyAspTyrVal-----ArgGlyArgLeuLysAsn 632
Db 1479 CCGTTAACGCATACG---GTCAGTTATGATTACGTTGATGCGCGTAATCGATTACCGAT 1535
QY 633 LeuProSerLeuProGlyArg 639
Db 1536 ACGCCA---TTACCCCGCGGT 1553

RESULT 36
US-09-307-973A-1
; Sequence 1, Application US/09307973A
; Patent No. 6432686
; GENERAL INFORMATION:
; APPLICANT: BULPHIUS, BEN
; APPLICANT: GATENBY, ANTHONY A.
; APPLICANT: TRIMBUR, DONALD E.
; APPLICANT: WHITED, GREGORY
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
; TITLE OF INVENTION: 1,3-PROPANEDIOL BY RECOMBINANT
; TITLE OF INVENTION: ORGANISMS COMPRISING GENES FOR
; TITLE OF INVENTION: VITAMIN B12 TRANSPORT
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 925 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/307,973A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,190
; FILING DATE: JUNE 30, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1245
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

```

i ANTI-SENSE: NO
US-09-307-973A-1

Alignment Scores:
Pred. No.: Length: 3,21e-08 Matches: 1845
Score: 175.00 Matches: 131
Percent Similarity: 31.87% Conservative: 80
Best Local Similarity: 19.79% Mismatches: 215
Query Match: 4.34% Indels: 236
DB: 4 Gaps: 30

US-09-936-377-2 (1-758) x US-09-307-973A-1 (1-1845)

QY	29	GlnSerValGlyLeuGluThrValThrValValGlyLys-----SerArgProArgAla 46
Db	61	CAGGATACCGCCGGGATCTCTCGTCGTACTTAAACCGTTTGAACAGCGCGCAGC 120
QY	47	ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysVallelleSerGlyAspThrLeu 66
Db	121	ACT-----GTGTTGCCAACCAACCAGTT-----GTACCGCGTCAGATATCGAC 165
QY	67	ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db	166	CGTGCGAGTCGACCTCGGTCAATGATGTCTCGCGCTCTCCGGCGGTCTCATATCAC 225
QY	87	GlnTyrglyGlyAla-----SerAlaProValilleArgGlyGlnThrGlyArg 104
Db	226	CAAAACGGCGGTTCAGCTCAGCTCATCTATTATTCGGGTCAACAATCCAGTCAT 285
QY	105	IleLysVal-----LeuAsnHisGisGlyGluThrGlyAspMetAla 118
Db	286	GTGTTGTTTAATGATGGCGTFACGCCTGAATCTCGCGGGGTGAGTGTCT---GCC 342
QY	119	AspPheSerProAspHisAlalleMetValAspThrAlaLeuSerGlnGlnValGlutle 138
Db	343	GACCTTAGC-----CAGTTCCTATTGGCTTGTCCAGCGTGTGAATAT 387
QY	139	LeuArgGlyProValThrLeuLeuTyrsrSerGlyAsnValalacGlyLeuValaspVal 158
Db	388	ATCGTGCGCGCGCTCCGTGTTTATGTTTCGATGCAATAGCGGGGTGGTGAATATC 447
QY	159	AlaaspGlyLysleProGluLysMetProGluasnGlyValserGlyGlyLeuGlyLeu 178
Db	448	ATCAGACGGCGC-----GATCAACCCGGAACG 474
QY	179	ArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyleasnlleGlyLeugly 198
Db	475	GAAATT-----TCAGAGGGTGGCGAAGCAATAGTTAT 507
QY	199	LysAsnPheValLeuHisThrGluGlyLeuTyrglySerGlyAspTyralaValPro 218
Db	508	CAGAACTATGATGTCCTACGCAC----- 531
QY	219	ArgTyrgAsnLeuLysArgLeuProaspSerProArgArgPhealaasnglyGlnHis 238
Db	532	-----CAACAACTGGGGGATAAGACACCGGTACCGTGTGGCGCATAT 576
QY	239	ArgAlaValLeuGlyTrpArgLysArgPheTyrgArggThrtyrSeraspArgasp 258
Db	576	----- 576
QY	259	GlnTyrglyLeuproAlaHisSerHisGluTyrgAspAspCyshisalaaspillelletrp 278
Db	577	-----GCCCATCTCATGGTTATGAT----- 597
QY	279	GlnLysSerLeuilleAsnLysArgTyrgLeuGlnLeuTyrgProHisLeuLeuThrGluGlu 298
Db	598	-----GTTGTTGGCTATGGTAAATACCGGAACGCAAGCG 630
QY	299	AspValaspTyrgAspAsnProGlyLeuSerCysglyPhe----- 311
Db	631	CAGACAGATAACGAT-----GGTTTTTAAGTAAAAACGGCTTTATGTC 672

```

Db      1516 GCGCGC 1521
RESULT 37
US-09-328-352-1816
; Sequence 1816, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1816
; LENGTH: 2286
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1816

Alignment Scores:
Pred. No.: 5,79e-08 Length: 2286
Score: 174.00 Matches: 181
Percent Similarity: 32.86% Conservative: 117
Best Local Similarity: 19.96% Mismatches: 299
Query Match: 4.31% Indels: 310
DB: 4 Gaps: 46

US-09-936-377-2 (1-758) x US-09-328-352-1816 (1-2286)

QY      7 LysProIleValLeuSerIleLeu-----LeuIleAsnThrProLeuLeuAla 22
Db      28 AAGCGTATTTCAGTCAGTGTCTTCTTTCAGTACTGCGAAGCATGATGCTATGGCT 87
QY      23 GlnAlaHisGluThrGluGlnSer----- 30
Db      88 TTTGCTGCACAAATGAGCAAGATCAAGCTGAAAGAAACATTAGAAAAGCGCTGTAACCT 147
QY      31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db      148 GTGAATTTGGAACAATTTTCGTACAGCTGAAGAGCAAGTGAAGCAATCGCTCGGTGTA 207
QY      51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
Db      208 -----TCGCTTATTACCAAGAGAGATTTAGAAAAAATACCA 243
QY      71 Val-----AsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyr 88
Db      244 GTCTGTAATGATATTCTGACTATATGTCGTATGCGACGAGTGTCAACCTGACA----- 297
QY      89 GlyGlyGlyAlaSerAlaProValIleArgGlyGlnThr----- 101
Db      298 -----GGCAATAGTGTACAGGGCAGCGTGTGTAATATAGACAAATGATATTCGCGGA 351
QY      102 -----GlyArgArgIleLysValLeuAsn--- 109
Db      352 ATGGGCGCTGAAACACGCTTATTTAGTGACGCGTAAACCGATTAATCTCGTAATCA 411
QY      110 ---HisHisGlyGluThrGlyAspMet-----AlaAspPheSerProAsp 123
Db      412 GTTCGTATGCTGCGAAGAGAGAGCGGTATACACGAGCGAGCTCAAACTGGGTACACGA 471
QY      124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db      472 GAAGCAATC-----GAGTCGATCGAAGTGTTCACGTTGGACCGACGA 510
QY      144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db      511 GCAGCTCGTATGGCTCTGGTCTCGGGTGTGTAGTTAATCATCTACTTAAAGTA 570
QY      164 -----ProGlyLysMetProGlu 169
          |||||

```

```

Db      571 ACRAATGAACCTCATGGTTTCAGTAGAGTCTATATCTTCACAGCCTGAAGACTCCAAAGAA 630
QY      170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSer----- 182
Db      631 ---GGTTCATCAAAATCGTGTGGTGTAAATGTAAGTGGCCATTAAATTAAGAGAGCTTTTG 687
QY      183 -----GlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIle----- 195
Db      688 TCTTATCGTTTATATGTTAATCAATAAACGGAAGCTGATGATGATGATATATAAA 747
QY      196 GlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys-----SerGly 213
Db      748 TCTATCGGTAGTACA---GCAGCTGGCGGTGAAGGTGTAAATAATAAGATATTTCCAGC 804
QY      214 -----AspTyrAlaValProArg 219
Db      805 CGTTTAGCTTGGCAGGCAACAGACCAAACTCTTCTGCGATATTTCTTCTAGCAAA 864
QY      220 TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
Db      865 CAAGGTAATATT-----TATTCTGGTGAATCTCAGTTAAATGCAAAATGCTGAAGCGGAT 918
QY      240 AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln 259
Db      919 GCGATTCITTC-----CAGCTGATGCTGAAGAAACCAATACCATCATGATCGTGATAGC 972
QY      260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGln 279
Db      973 TATGCATTA-----ACGCAGAA-----GGTGATTGGTCTTGGGT 1008
QY      280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluAsp 299
Db      1009 AAGAGTAAGTATGTTGCTCAATATGATGAAGACCAATCAACAACTGCTACCTGAA----- 1062
QY      300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAsp-----Asp 314
Db      1063 -----GGCTTGGCGGAAGTGTAGAAAGGAAAAATTAATAATCTTGAT 1104
QY      315 AspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuAspAsnLysArg 334
Db      1105 GATAAAGCCATCTCG-----CGTTTGAAGAACT 1131
QY      335 TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPhe-----GluAlaLeuArg 352
Db      1132 CTTTCGCTTAAACGCGAGGCTAATATCTCTTTGAATACTATTTACCCCAAGTATTAAC 1191
QY      353 ValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAsn 372
Db      1192 GTA-----GGTACCGAATGGGTTGAAGAC 1215
QY      373 PhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArg 392
Db      1216 AGNTTTAAAGATAATGCTCGACAACT-----CAA 1245
QY      393 LeuLysGlySerTyrGlyValGlnTyr-----LeuGlyGlnLysSerSer 407
Db      1246 GGTAAAGACAGCAGCGTGTTCAGGTTTACGATCAATTTAGCGAAGGTTGATCGTAGTAA 1305
QY      408 AlaLeuSerAlaThrSerGluAlaVal-----LysGln 418
Db      1306 ATGGAGTCACTGCTTCTGCTATATATTGAGATAACCTGAAAGTTACAGACAGCA 1365
QY      419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
Db      1366 GATGTGTATTAGGTTTACGTTTGTGATGACCATAGTAAATCTGGT-----TCTAAT 1416
QY      439 Trp-----AspAsnPheThrLeuGluGlyCly 447
Db      1417 TGGAGCCCAAGCTTAAATATTACTCAAAAACCTCAATGATTTTCACTTTTAAAGGTGG 1476
QY      448 ValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGlu 467
Db      1477 GTAGCA-----AAAGCT----- 1488

```


QY 468 AsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaAspSerPhe 487
 Db 1489 -----TATAAAGCACCA-----AATATGTATCAAAATGCCGAAGGTAT 1527
 QY 488 AlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis 507
 Db 1528 TTATTAAAGTACAAATGGCAATGGCTGTCTCT----- 1557
 QY 508 GlnGluArgLeuProSerThrGlnGlnLeuTyrAlaHisGlyLysHisValAlaThrAsn 527
 Db 1558 -----GCTAATATTAGTCTCGGT 1575
 QY 528 ThrPheGluValGlyAsnLysHisLysAsnLysGluArgSerAsnAsnLysLeuAla 547
 Db 1576 TGTATTATACAGGTAAATCGTATTAAACCTGAAACATCGTAAACAAAGAGCTCGT 1635
 QY 548 LeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
 Db 1636 ATTCAAGTCCAAAGAGATATCGTAATCGAGCTTAACCTGCTCCGTAATGATTATAA 1695
 QY 568 AsnTyrIleTyrAlaGlnThr-----LeuAsnAspGlyArgGlyProLysSer 583
 Db 1696 GATAAGATTGTCGGGTACTCATGTTGCGAACAGTTGATGGCTCAAGTACAAATGCA 1755
 QY 584 -----IleGluAspAspSerGluMetLysLeuValArgTyr---AsnGlnSerGly 599
 Db 1756 AATACAGGAGCTGTGACCAATACGAAGTGAATATTTTCGTTGGAAATAATACGCTAAA 1815
 QY 600 AlaAspPheTyrGlyAlaGlnGlyGluIleTyrPheLysProThrProArgTyrArgIle 619
 Db 1816 GCCTTAATCAAGGTTTTCAGGAAGT-----TTG 1845
 QY 620 GlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
 Db 1846 GGGTTAGACTTCGGTGATATCCGCTGG---ACTAATAACTTTACCTACATGATGACTCG 1902
 QY 640 GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
 Db 1903 AAAGACAAGCAACACGGGAACCATTA-----TCTTTAGTT 1938
 QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
 Db 1939 CCAATCTATACAAATTAATCAATTTTGTGATTATGACATTAATGATCAATTTGATGTAAT 1998
 QY 680 LeuAspTyr-----TyrArgValPheAlaGlnAsnLysLeu--- 691
 Db 1999 TTTGTTATTACTCAATATGCTCGTGCATAATCAGTCAATTTGCAGAAATAGACTTAA 2058
 QY 692 -----AlaArgTyr 694
 Db 2059 TCCGTTATAGTTTCAGGAGTGCAGAAATCTCGCTTAAGCCAAAGTACGTAAGATTAT 2118
 QY 695 GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyr-----ArgArg 712
 Db 2119 AGTACCGCTGGT-----ATTAGTTGGTTATTAAGTTTCAGACCAAAAT 2163
 QY 713 AsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLysAsnSer 732
 Db 2164 AGTACACGCTGTGCT-----GTGAGTAATCTGTTTGATAAACA 2202
 QY 733 ValTyrAlaHisSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThr 751
 Db 2203 ATTTTAAAGACAGATTAATCTATTAGCCAAATATATATGACCCAGGTGAGCTTATTAC 2262
 QY 752 GlyGlyValAsnValLysPhe 758
 Db 2263 GCATCTTTAAATAATATCTTTC 2283

RESULT 38

US-09-453-702B-174

; Sequence 174, Application US/09453702B

; Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
 Burland, Valerie
 Perna, Nicole T.
 Plunkett, Guy
 Welch, Rod
 TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 NUMBER OF SEQUENCES: 265
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Quarles & Brady
 STREET: 1 South Pinckney Street
 CITY: Madison
 STATE: WI
 COUNTRY: US
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/453,702B
 Filing DATE: 03-Dec-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/110,955
 Filing DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27396
 REFERENCE/DOCKET NUMBER: 960296.95017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 251-5000
 TELEFAX: (608) 251-9166
 INFORMATION FOR SEQ ID NO: 174:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7304
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 174:

US-09-453-702B-174

Alignment Scores:
 Pred. No.: 3,89e-07 Length: 7304
 Score: 174.00 Matches: 171
 Percent Similarity: 31.41% Conservative: 101
 Best Local Similarity: 19.75% Mismatches: 288
 Query Match: 4.31% Indels: 306
 DB: 4 Gaps: 42

US-09-936-377-2 (1-758) x US-09-453-702B-174 (1-7304)

Qy 1 MetAlaGlnThrThr-----LeuLysProIleValLeuSerIleLeuLeu 15
 Db 5095 ATGGCTAAGTTCACACTTCATCTCAGGAATCAAGGTGCGCGCTCTTTTCACTGCTC 5154
 Qy 16 IleAsnThrProLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThr 35
 Db 5155 TTTCGACGACCAATGATT---CATGCAACCGACATCGAACGACCAAGATGGCGAAACA 5211
 Qy 36 ValThrValVal-----GlyLysSerArgProArgAlaThrSerGlyLeu-----Ileu 51
 Db 5212 ATCATCTGTTACAGCGGATGCAAAATACCGCACTGAGCGGACCGATGTTATCAACTCTG 5271
 Qy 52 HisThrSerThrAla----- 56
 Db 5272 AGCACTTCCACGCGGACATTAAACGATATGCCGATGCTGGATATCCCGAGGTGCTCAAT 5331
 Qy 57 -----SerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGly 74
 Db 5332 ACGGTTAGCGATCAGGTTCTCGAA-----AATCAGAATGCAACGACGCTGGAT 5379

QY	75	AspAlaLeuAspGlyValProGly--ileHisAlaserGlnTyrGlyGlyAlaSer	93
Db	5380	GAAGCGCTTATTAACCTGAGTAACGCTACAGCAACATATAGCGCACTCAGGAC	5439
QY	94	AlaProValileArgGlyGlnThrGlyArgArgileysValLeuAsnHisHisGlyGlu	113
Db	5440	GCCTTTGACGTCGTGTTTGTGCTAACCGG-	5472
QY	114	ThrGlyAspMetAlaAspPheSerProAspHisAlaileMetValAsp	129
Db	5473	-----GATGGCTCCATCATGACCAACGCGCTCGGAAC	5505
QY	130	-----ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal	143
Db	5506	GTACTTCCTCGCAGTTTCAACGCGGCACAGAACGTGTGAAGTCTTAAGGTCGCGCC	5565
QY	144	ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysile	163
Db	5566	TCCACGCTGTATGGCATCTCTCGATCTCTGTGTGATGATTAACTCGTGACC--AAGCGC	5622
QY	164	ProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly	183
Db	5623	CCGGAATAAAATTCATGCTTCTGTCTAGCC-	5655
QY	184	AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLys	199
Db	5656	-----ACGTCCTCCAGTTTGGCGGAGGCATCGGCAACTGATATCACA	5700
QY	200	AsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArg	219
Db	5701	GGTCCCAATGAAGCACTCAGCTGGCATACCGCTGACGGGGAAGTG--	5748
QY	220	TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg	239
Db	5748	-----	5748
QY	240	AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln	259
Db	5749	-----CAGGATGAAGATTACTGGCGAAATTCGCTAAGAGCGCAGTACA	5793
QY	260	TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGln	279
Db	5794	TTTATTGCCCGCTCA-----LeuLeuThrGluGluAspValAsp--CTCACTGGTTT	5820
QY	280	LysSerLeuLeuAsnLysArgTyrLeuGlnLeuTyrProHis	293
Db	5821	GGTGAT-----AATGCAACAGTAACCATGCTCTATTCCCATCGGCACTATAAACTCCG	5874
QY	294	-----LeuLeuThrGluGluAspValAsp--	301
Db	5875	TTCGATCGTGGACGATTTTCGACCTTACAGCAACAGACGCCGTAAACGTTGATCGAATA	5934
QY	302	-----TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla	319
Db	5935	ATACGTTTTTCACGAACCG--TTTAATATTACAGAT--	5967
QY	320	HisAlaHisAsnGlyLysProTyrPileAspLeuArgAsnLysArgTyrGluLeuArgAla	339
Db	5968	-----GCTAGTCGATCTCGGCCAATCAACGAGCAATATCATCTCAATAGC	6015
QY	340	GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp	359
Db	6016	CAGTGGACAGCGCGCTTT-----GAT	6036
QY	360	TyrHisHisaspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGln-	377
Db	6037	TACAGCTACAGCCAG-----GATAAATACAGCGCAATCATCGCTCGC	6078
QY	378	-----ThrGlnAsnAlaArgIleGluLeu-	385
Db	6079	GTTTACCGCGTATGATGCAACGACAGGACCGCTGACACGGCGTGTGTGATGCACTCAGGGA	6138

386	-----ArgHisGlnProIleGlyArgLeuLeuGlySerTyrPheGlyValGlnTyr	401
6139	TCACCCAGCGGTATGCATCTTACTCGTGGCGGATCTTCAAGGGAAT-----GTTGATATT	6192
402	LeuGlyGlnLysSerSerAlaLeuSerAlaThrSer-----GluAlaValLys	417
6193	GCTGGGTCTATATAGATTCTGGGTGGGTCTCATATGAATATATGATCTTCGCG	6252
418	GlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAla	437
6253	ACAGATATGATTCTGCTGTAATAAAGCTAAAGATTTCATATCTACAACCCCGTTTATGGC	6312
438	AsnTrpAspAsnPheThrLeu-----GluGlyValArgValGluLysGln	453
6313	AAATACCAAAATGCACACCGTTTCGGCTCGACAGTGCATCAGACGATCAACACAGAG	6372
454	LysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsn-----	468
6373	AGCTACTCAGCTTATGCACAGGATGGCTCTATCTGACCGATAAATGGATTGCGCTGCC	6432
469	-----TyrTyr-----	470
6433	GGGATCCGCTATCAGTATTATACACGAGTATGCGGGTAAGGCCGTCCTTTTAATGTCAT	6492
471	-----LysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg	485
6493	ACTGACAGCGCGATGAACAATGGACGCCCAAATCGGGTTAGTCTACAAACTGACGCCA	6552
486	SerPheAlaLeuSerGlyAsnTrp-----TyrPheThrProGlnHisLysLeuSerLeu	503
6553	TCGGTATCCTTATTGTTAAATTTATTCGAAACATTTATGCCCCAA-----TCGTC	6603
504	ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis	523
6604	ATTGCCAGTACATCGGAGATCTTCACCGGAA-----	6636
524	ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn	543
6637	---TCATCTAATGCTTACGAATCGGGCGCAAAA-----	6666
544	IleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg	563
6667	TTGAGCTTTCGATGGTATACCGCAT-----ATTGCGCTGTTGAT	6711
564	AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----	577
6712	ATCCATAAACGTAACAGCTGTGTATACCGAAAGATTTCGTGATCAAAACCATGCCAAAACG	6771
578	---GlyArg--GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr	595
6772	GCAGCGCGGTCGTTCAAGAGGGGTAGAAGTCGACTTCGCGGAGCATTAATCTGAAAC	6831
596	AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro	615
6832	ATTAATATCAITTCGACGTACCGGTATACCGATGCTAAGGTTCTCGAGATCCT-----	6885
616	ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer	635
6886	-----GANTATGCGAGGAAACCATTCGGAATGTTCTCTCGT	6921
636	LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspGlnAsn	655
6922	CATACCGGT-----TCGCTATTCTCGACCTATGACATTCATTAAC	6960
656	AlaProArgValProAlaAlaArgLeu-----GlyPheHisLeu-----LysAlaSer	671
6961	ATGCCAGCAATAACACACACTGACGCTTTGGCGGTGGTGGACATTTGGGTAAAGCGCTCG	7020
672	LeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeu	691
7021	GCACCAAT-----GGGCGTGCATTTAT-----	7044
692	AlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg	711

Db 7045 -----CTGCCAGGCTATTTCTGGCGGATGCTTCCCGCATACAAA 7086
 Qy 712 ArgAsnThrArgTyrGlyGluTrpAsnTyrValLysAlaAspAsnLeuLeuAsnGln 731
 Db 7087 ATGAATTGAGTAT---CCGGTCACACTGCAATTAAACGTCAAAAACCTGTTTGATAA 7143
 Qy 732 SerValTyrAlaHisSer 737
 Db 7144 ACGTAATTACACTCTTCC 7161

RESULT 39

US-09-252-991A-1708
 ; Sequence 1708, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 1708
 ; LENGTH: 2172
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-1708

Alignment Scores:

Pred. No.: 6,01e-08 Length: 2172
 Score: 173.50 Matches: 181
 Percent Similarity: 33.89% Conservative: 103
 Best Local Similarity: 21.60% Mismatches: 285
 Query Match: 4.30% Indels: 272
 DB: 4 Gaps: 44

US-09-936-377-2 (1-758) x US-09-252-991A-1708 (1-2172)

Qy 7 LysProIleValLeuSerIle-----LeuLeuIleAsnThrProLeuLeuAlaGln 23
 Db 40 GAGCTTTTCGCTTCGCTCTCCACCGCTCTGCTGTACCGCGCTCGCCGAA 99
 Qy 24 AlaHisGluThrGluGlnSerValGlyLeuGluThrValThrVal----- 38
 Db 100 ACCGAGGAGCGCGCTCGCCCTCGCCCGAGCCAGTGTGTCAGCGTTGCGCAGGATCCCGCC 159
 Qy 39 -----ValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSer 54
 Db 160 GAACCTCGACCATCGACTCGCCACTCCGTCAGTCCGGTTCGGCTTCGGCTCAGC 219
 Qy 55 Thr-----AlaSerAspLysIleSerGlyAspThrLeuArg----- 67
 Db 220 GCCCTGGACACCCCGCGGAGCAGCAGCATCAGCGCGAGGAGTTCGCGGAGCAAC 279
 Qy 68 -----GlnLysAlaValAsn-----LeuGlyAspAla 76
 Db 280 AACCCAGCGTCGAGCGCGCTGACCCGCGAGCCCGGATCAGCTTCATCGGCCCA 339
 Qy 77 LeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal 96
 Db 340 GGATGCGCGTACCGCGCTTCGCGCGCGCTTCAGCGCGGATTCGCGCGGATTCGCGCGG 393
 Qy 97 IleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAsp 116
 Db 394 ATGCACTGTCGAGCGCAGCGCGCTACACCG-CATGGCAGCGTGAATT----- 446
 Qy 117 MetAlaAspPheSerProAspHisAlaIle-MetValAspThrAlaLeuSerGlnVal 136
 Db -----GlnGlnVal----- 136

Db 447 -----CCCAGGCGATCCGTGGATGGTTC-----GAGCGCAT 476
 Qy 136 lGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVa 156
 Db 477 CGAGCTGATCCGCGCGCGCGCTCGGTGCTGTACGCGAAGCGCCACCGCGCGGTGAT 536
 Qy 156 lAspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLe 176
 Db 537 CAACGTG-----GTGCCGAAGAAGCCCTTCGCCGCGGAGAT 572
 Qy 176 u-----GlyLeuArgLeuSerSerGlyAsnLeuGluLys-----LeuThrSe 190
 Db 573 CCGCAACCACTCGCTCGCTACGCTCTCTACGACCAACCGCCAGTTCGCCCTCGACAG 632
 Qy 190 rGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu---Ty 209
 Db 633 CGCGCGC-----TCGCTGACCGACAGCAGCTCAGCTA 662
 Qy 209 rArgLysSerGlyAspTyrAlaVal---ProArgTyrArgAsnLeuLysArgLeuProAs 228
 Db 663 CGGCTCAACCT-GAACCAGCAGCAGCAGCCGCGCTG-----GA 700
 Qy 228 pSerProArgArgPheAlaAsnGlyGlnHisArg-AlaValLeuGlyTyrArgLysArg- 247
 Db 701 TCACCGTGGCGACTCGCGCAACCTCGGCATCAGCGCGCGCTGCGCTGCGAGGCCAGCG 760
 Qy 248 -----PheTyrArgArgThrTyrSerAspArgArg-----AspGlnTyrG 261
 Db 761 ACGATCTGGCTTCACCTCGCCACGACTATGGCGACCAGGAGCGGATGAACGACTTCG 820
 Qy 261 lLeuProAlaHisSerHisGlyTyrAspCysHisAlaSerIleIleTrpGlnLys 281
 Db 821 GCACCCGCTGCTCGCGCGCAAGTAC-----CACAGC 853
 Qy 281 erLeuIleAsnLysArgTyr-----LeuGlnLeuTyrProHisLeuLeuT 296
 Db 854 GCCTGCGCGAGAGAACTCAACGTCGCAACGCTGCGAGCGCTACACGACCGAGTGA 913
 Qy 296 hrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspA 316
 Db 914 CCGCTCTGACCGAGCGACTGG-----AGCCTTTCC-----G 943
 Qy 316 spAlaHisAlaHisAlaHisAsnGlyLysProIleAspLeuArgAsnLysArgTyrG 336
 Db 944 ACAGC-----GTACCCGCGAGCAACAGTGTACT 973
 Qy 336 lueuArgAla-----GluTyrLysGlnProPheP 346
 Db 974 ACATCAAGCGCCGCGCCACTGCGCAACGCGGAAACCTACGAATGGGACGCTCCCG- 1029
 Qy 346 roGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr-----His 362
 Db 1030 -----CGCAAGAGCTGTTCGCGAGGAGTACCTGCGCATCAGCCAGC 1072
 Qy 362 isAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAla 382
 Db 1073 AGCAGGAGCAGATCGCGCAGCCGCGAGACTTCGCTTC-----CAGCAGCGCC 1120
 Qy 382 rGleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrL 402
 Db 1121 T-GTTCGCTCGACCGCCACCTGTCGCGCGCA----- 1157
 Qy 402 euGlyGlnLysSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuL 422
 Db 1158 -----GTCAACCGCATCCGC 1173
 Qy 422 euAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsn 442
 Db 1174 TTCGCGCTGACCAACACATATCGCCCTATAC----- 1205
 Qy 442 heThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysA 462
 Db -----CGATGTCGCGCGCGACTACATCGAC 1230

```

QY 462 laLeuileaspArgGluAsnTyrTyrLysGlnProLeu-ProAspLeuGlyAlaHisArg 481
Db 1231 CCTGGCATCCCGCGCGCGCTACTTGAAGCGGTTCGCCCTACCGCGCGCACTCGCGC 1290
QY 482 GlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeu 501
Db 1291 AGCCAGACCGCGACCTTCGCCCTGTTCGCCGAGAACCGCTGCAGCTCAACGAGCGCTG 1350
QY 502 Ser-----LeuThr 504
Db 1351 TCGCTGGTACCGGGGTGCGCGCGACGACCATATACCGCGACGACCTGCTCAC 1410
QY 505 AlaSerHisGlnLysArg-----LeuPro 512
Db 1411 GGCACCGCGACGCGCGCGCTGCGCGCGCACTGCGCGCGCGCGGATGCTTCGCG 1470
QY 513 SerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn-----527
Db 1471 CTGACCGCGGAAGTGTGCTGTACGCGCGAGTACTCCACCGAGCGAGCGGTGTGAGCAAC 1530
QY 528 -----ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnLysGlu 545
Db 1531 CTGATCACCTCAACCGCGCGCGCGAGAGTGGACCTGACCCACTGCAAGCAGACCGAG 1590
QY 546 LeuAlaLeuGly-----TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg 563
Db 1591 GTCGGCTCAAGCAGCTGTTCGCGCGGTGCGCGCGGAGTGGACCTGCGCGCTAC---1647
QY 564 AsnArgPheGlyAsnTyrLysTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer 583
Db 1648 -----CATATGCTCAAGAAAGAGTGTCTCAGCGCC---AACCCTGCGCG 1689
QY 584 IleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 1690 CCGCAGCAGCGCGCAGAGTTC-----GGCCAGCAGAGTTCGAGC-----1728
QY 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 1729 GGCCTGGAGCGCAGCTGGAACCTGGAACCTGGCCAGGAGTGGCGGTGTTCGCGAGCGC 1788
QY 624 AspTyrValArgLysLeuLysAsnLeu-----ProSerLeuProGlyArgGlu---Asp 641
Db 1789 GCGCTGGTGTGCGCGAATACGACGACTTCGACGAAACGATGACGCGGACGACTACTCG 1848
QY 642 AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
Db 1849 CGCAACGGCAATCGCCCC-----AGGAAGTGCACCG---CGCAC 1887
QY 662 AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
Db 1888 GCCAACCTTGTGCTGGACAAAGTCTTCGCGGAGACCTCGCGGTGCGCGCGGTGCGC 1947
QY 682 TyrTyr---ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db 1948 TACGTGACCGCGCTACGCGAC-----GGCGCAACACGCGCCAGCCTGCGCGCGC 1998
QY 701 HisHisMetLeu-----AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu 718
Db 1999 TACACGGTGTGCTCAACTCGGC-----2025
QY 719 TrpAsnTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyr 734
Db 2026 -----TGCGGGTACGGCGGACCTGACCTCGCGCTTGAGCTGTAT 2067

```

RESULT 40
 US-09-252-991A-2016/c
 ; Sequence 2016, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2016
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-2016

Alignment Scores:
Pred. No.: 6,32e-08 Length: 2241
Score: 173.50 Matches: 181
Percent Similarity: 33.89% Conservative: 103
Best Local Similarity: 21.60% Mismatches: 285
Query Match: 4.30% Indels: 272
DB: 4 Gaps: 44

US-09-936-377-2 (1-758) x US-09-252-991A-2016 (1-2241)

QY 7 LysProIleValLeuSerIle-----LeuLeuIleAsnThrProLeuLeuAlaGln 23
Db 2196 GAGCCTTCGCGCTCGGCTCTCCACAGCCTGCTGTGTACCGCGCGCTCGCGGAA 2137
QY 24 AlaHisGluThrGluGlnSerValGlyLeuGluThrValThrVal-----38
Db 2136 ACCGAGGAAGCGCGCTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2077
QY 39 -----ValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSer 54
Db 2076 GAACCTCGACCATACGACCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCG 2017
QY 55 Thr-----AlaSerAspLysIleIleSerGlyAspThrLeuArg-----67
Db 2016 GGCCTGGACACCGCGCGGAGCACCAGCATCATCAGCGCGGAGGAGTTCGCGACGCAAC 1957
QY 68 -----GlnLysAlaValAsn-----LeuGlyAspAla 76
Db 1956 AACCCAGCGTCCAGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1897
QY 77 LeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal 96
Db 1896 GGAGATGGCGGTACCGCGCTCTCGCGCGCGCGCTTCAGCGCGCGCGCGCGCGCG 1843
QY 97 IleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAsp 116
Db 1842 ATGCAACTGTCGACGCGCGCGCGCTGTACACCGG-CATGGGCGCGCGTGAACCT-1790
QY 117 MetAlaAspPheSerProAspHisAlaIle-MetValAspThrAlaLeuSerGlnGlnVa 136
Db 1789 -----CCCCAGCATCCGTGATGTC-----GAGCGCAT 1760
QY 136 IclulleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVa 156
Db 1759 CGACGTGATCGCGCGCGCGCGCTCGCGCTGTACGCGGAAAGCGCGCGCGCGCGGTAT 1700
QY 156 IaspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLe 176
Db 1699 CACCTG-----GTGCGGAGAGAGCGCTTTTCGCGCGCGAGAT 1664
QY 176 u-----GlyLeuArgLeuSerSerGlyAsnLeuGluLys-----LeuThrSe 190
Db 1663 CGGCAACCACTCGCGCTCGCGCTGTACGACCGCGCGCGCGCGCGCGCGCGCGCG 1604
QY 190 rGlyGlyIleAsnLysGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu---Ty 209
Db 1603 CGCGCGC-----TCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1574
QY 209 rArgLysSerGlyAspTyrAlaVal---ProArgTyrArgAsnLeuLysArgLeuProAs 228

```

```

Db      1573  CCGGCTCAAGCT-GAACCAGCAGCAGCAGCCAGCGCTG-----GA 1536
Qy      228  pSerProArgArgPheAlaAsnGlyGlnHisArg-AlaValLeuGlyTrpArgLysArg- 247
Db      1535  TCGACCTGGCGACTCGGCAACTGGGATATAGCGGGCGCTGCGCTGGCAGCCAGCG 1476
Qy      248  -----PheTyrArgArgThrTyrSerArgArg-----AspGlnTyrG 261
Db      1475  ACGATCTGGCCTTCCACCTCGCCACAGCTATGGCAGCAGGAGCGATGAACAGCTTCG 1416
Qy      261  lyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleTyrGlnLys 281
Db      1415  GCACCCGCTGGTGGCGGCAAGTAC-----CACAAAGC 1383
Qy      281  erLeuIleAsnLysArgTyr-----LeuGlnLeuTyrProHisLeuLeuT 296
Db      1382  GCCTGGCGGAGAAACTTACAACGTGCGCAACGAGTGCAGCGCTTACAACAGCAGTGA 1323
Qy      296  hrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspA 316
Db      1322  CCGCCTGACACGAGCTGG-----AGCCTTTC-----G 1293
Qy      316  spAlaHisAlaHisAlaHisGlnGlyLysProTrpIleAspLeuArgAsnLysArgTyrG 336
Db      1292  ACAGC-----GTACCGGAGCAACCACTTGTACT 1263
Qy      336  luLeuArgAla-----GluTrpLysGlnProPheP 346
Db      1262  ACATCAAGCGCCGCGCCACTGGCGCAACGCCGAACCTACGAATGGAGTCCCG----- 1207
Qy      346  roGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr-----HisH 362
Db      1206  -----CGCAAGAGCTGTTGGCAGGAGCTACTTGGCGATCAGGCACG 1164
Qy      362  leAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThrGlnAsnAlaA 382
Db      1163  AGCAGAGCATCGGCGACCGCAGACCTTCGCTTC-----CAGCAGCC 1116
Qy      382  rgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrL 402
Db      1115  T-GTTCGGCCTCGACAGCGCAGCCCTGCTGGCGCCGA----- 1079
Qy      402  euGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuL 422
Db      1078  -----GTACAAACCGCATCCGC 1063
Qy      422  euAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsnP 442
Db      1062  TTCCGCTTGACCAACACATATCGCCCTATAC----- 1031
Qy      442  heThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysA 462
Db      1030  -----CGATGTCGGCGGCTACATCGAC 1006
Qy      462  laLeuIleAspArgGluAsnTyrTyrLysGlnProLeu-ProAspLeuGlyAlaHisArg 481
Db      1005  CCCTGGCATCCGCGCCCGGCTACTTTCGAAAGCGCTGCGCCTACCGCGCGCATCGCGC 946
Qy      482  GlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu 501
Db      945  AGCCAGACCCGACCTTCGCTCTGTCGCCGAGAACCGCCTGCAGCTCAACGAGCGCCTG 886
Qy      502  Ser-----LeuThr 504
Db      885  TCGCTGGTGACCGGGGTGGCGCGCAGCAACCATATCGACCGCGAGCAGCTGCTCACC 826
Qy      505  AlaSerHisGlnGluArg-----LeuPro 512
Db      825  GGCACCCGCGAGCGCGAGCGCTGCGAGGGGCAACTGCGCGCGGAGTGGTCTTCGCG 766
Qy      513  SerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn----- 527

```

```

Db      765  CTGACCCCGGAACTGTGCTGTACGGCCACTCTCCACAGCAGCAGGACGGTGTGCAAC 706
Qy      528  -----ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu 545
Db      705  CTGATCACCTCAACGGCGGCCAGCAGCATGGACCTGACCCACTCGAAGCAGACCGAG 646
Qy      546  LeuAlaLeuGly-----TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg 563
Db      645  GTCGCGCTCAAGCAGCTGTTCCCGAGCGGTCCGGCGAATAGGACCTGGCGCGCTAC--- 589
Qy      564  AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer 583
Db      588  -----CATATCGTCAAGAAGAGCTGCTCAGCGCC---AACCCGCTGCCG 547
Qy      594  IleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db      546  CCGCAGCAGCGCGCAGCGAGTGC-----GGCCAGCAGAGTTCGGAC----- 508
Qy      604  GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db      507  GGCCTGGAGGCCAGCTTGAACCTGCGCCAGGACTGGCGGTCTCGCGCAAGCC 448
Qy      624  AspTyrValArgGlyArgLeuLysAsnLeu---ProSerLeuProGlyArgGlu---Asp 641
Db      447  GCGCTGGTGGTGGCGAATACGACGACTTCGAGCAACGATCGACGGCGACCTACTCG 388
Qy      642  AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
Db      387  CGCAACGGCAATCGCCCC-----AGGAACGTGCCACCG---CGCACC 349
Qy      662  AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
Db      348  GCCAACCTCTGGCTGGGACAGTCTCTCGCGAGACCTGCGCGCTCGCGCGCTGCGC 289
Qy      682  TyrTyr---ArgValPheAlaGlnAsnLysLeuAlaArgTyrGlyThrArgThrProGly 700
Db      288  TAGCTCGACCGCGCTAGCGCGAC-----GCCGCCAACAGCGCCAGCTCCCGCGC 238
Qy      701  HisHisMetLeu-----AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu 718
Db      237  TACACGGTGGTGGACTCCCAACCTGGGC----- 211
Qy      719  TrpAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyr 734
Db      210  -----TGGCGGTACGGCGGACCTGACCTCGGCTCGGCTGAGCTGTAC 169

```

Search completed: November 17, 2003, 19:58:13
Job time : 1837 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 19:22:57 ; Search time 461 Seconds

(without alignments)
5373.750 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLILLINTEL.....FLSDTFQMGSRFTGVNVKF 758

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09936377/runat_14112003_104451_22224/app_query.fasta_1.903
-DB=Published Applications NA -QFMT=fastap -SURFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blousm62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09936377@cgn_1_154@runat_14112003_104451_22224
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Query Match	Length	DB ID	Description
1	3328	97.3	2295	14 US-10-066-551-7

2	230	5.7	1830121	14	US-10-329-960-1	Sequence 1, Appli
3	227.5	5.6	2307	14	US-10-066-551-8	Sequence 8, Appli
4	223	5.5	2381	11	US-09-801-451A-3	Sequence 3, Appli
5	220.5	5.5	32159	12	US-10-085-959-54	Sequence 54, Appli
6	207	5.1	2600	11	US-09-801-451A-1	Sequence 1, Appli
7	198.5	4.9	2061	12	US-10-181-319-1	Sequence 194, App
8	188	4.7	9057	14	US-10-114-170-194	Sequence 731, App
9	185	4.6	1980	12	US-10-238-075-731	Sequence 172, App
10	185	4.6	9054	12	US-10-085-959-172	Sequence 507, App
11	184	4.6	12264	12	US-10-238-075-507	Sequence 11, Appl
12	184	4.6	23654	12	US-10-085-959-11	Sequence 526, App
13	181	4.5	1962	12	US-10-238-075-526	Sequence 4, Appli
14	180.5	4.5	5099	14	US-10-043-344-4	Sequence 4030, Ap
15	178.5	4.4	2442	9	US-09-815-242-4030	Sequence 262, App
16	177.5	4.4	2139	12	US-10-238-075-262	Sequence 152, App
17	177.5	4.4	7306	12	US-10-085-959-152	Sequence 261, App
18	177.5	4.4	7315	12	US-10-238-075-261	Sequence 24, Appl
19	176.5	4.4	6732	12	US-10-085-959-24	Sequence 174, App
20	174	4.3	7304	14	US-10-114-170-174	Sequence 13, Appl
21	169.5	4.2	25801	12	US-10-181-319-13	Sequence 16, Appl
22	164.5	4.1	6737	14	US-10-114-170-76	Sequence 109, App
23	164.5	4.1	76804	12	US-10-085-959-109	Sequence 3, Appli
24	161.5	4.0	5009	14	US-10-043-344-3	Sequence 2, Appli
25	161.5	4.0	5033	14	US-10-043-344-2	Sequence 119, App
26	160.5	4.0	3275	14	US-10-114-170-119	Sequence 719, App
27	160.5	4.0	5885	12	US-10-238-075-719	Sequence 59, Appl
28	160.5	4.0	8249	12	US-10-085-959-58	Sequence 1, Appli
29	160.5	4.0	1830121	14	US-10-329-960-1	Sequence 1076, Ap
30	160	4.0	1620	12	US-10-238-075-720	Sequence 5, Appli
31	159	3.9	87563	14	US-10-114-170-57	Sequence 1049, Ap
32	156	3.9	6376	14	US-10-098-808-1	Sequence 1, Appli
33	155.5	3.9	2046	12	US-10-238-075-1076	Sequence 1076, Ap
34	155.5	3.9	2748	12	US-10-240-218-5	Sequence 5, Appli
35	155.5	3.9	12943	12	US-10-238-075-1049	Sequence 1049, Ap
36	155	3.8	4699	14	US-10-043-344-1	Sequence 1, Appli
37	154	3.8	5144	14	US-10-043-344-105	Sequence 105, App
38	150.5	3.7	3286	9	US-09-332-226-1	Sequence 1, Appli
39	148	3.7	3537	9	US-09-332-226-3	Sequence 3, Appli
40	141.5	3.5	2190	9	US-09-815-242-6035	Sequence 6035, Ap
41	135	3.3	2481	10	US-09-881-752A-333	Sequence 333, App
42	131	3.2	9025608	14	US-10-156-761-1	Sequence 1, Appli
43	129	3.2	36321	14	US-10-187-267A-1	Sequence 4051, Ap
44	127	3.1	2448	9	US-09-815-242-4051	Sequence 284, App
45	127	3.1	7092	14	US-10-156-761-284	

ALIGNMENTS

RESULT 1

US-10-066-551-7
; Sequence 7, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 7
; LENGTH: 2295
; TYPE: DNA


```

; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)
Alignment Scores:
Pred. No.: 2,87e-10 Length: 1830121
Score: 230.00 Matches: 171
Percent Similarity: 36.63% Conservative: 125
Best Local Similarity: 21.16% Mismatches: 285
Query Match: 5.70% Indels: 229
DB: 14 Gaps: 41

US-09-936-377-2 (1-758) x US-10-329-960-1 (1-1830121)
Qy 11 LeuSerileLeuLeuleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
Db 291699 CTTTCCTCGCAATTCACACACC-----TTAGTGACAGCAAAATGCGCTAGCGCAATCC 291752
Qy 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db 291753 GTTGAAATTAGACTCTATCAACGTTATTGCGACACGAGAYCCA-----AGTAGGTTT 291803
Qy 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
Db 291804 GCTTATAGCGCAGAAAAACAATCTAAA-----GATAGTCTTCTTCTAGCAAGCG 291854
Qy 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyGlyGly 90
Db 291855 ACTAGTGTTCAGATGCGTTAGAACACATCCCAATGTTGATGTAGA-----GGCGGT 291908
Qy 91 Gly-----AlaSerAlaProValIleArgGlyGlnThrGlyArgArg-----IleLys 106
Db 291909 TCGAGAAGCATTTGCTCAAAACCTAATATTTCGAGGGTTAGTGATATCGTGTGTGCAA 291968
Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
Db 291969 GTCATTGATGCGGTGACACAAAATTTTGATTAGACATAGAGT-----TCTTAT 292019
Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeu 146
Db 292020 TTTCTTCCAATGTCACTCATCCAGAAATTAAGATTAATCAAGGACCAAGTAGTCCCTTA 292079
Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166
Db 292080 TGGGGTAGCGGTGCATTTGGGTGGTGGTGGCAATGCGTAGCCAAATGCTTTAGACTTA 292139
Qy 167 MetProGluAsn-----GlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly 183
Db 292140 TTGAAAAAATAATGACAAATTCGAGTTAAAAATTCGCCAAGGTTATCAAACTGCTAATAAT 292199
Qy 184 AsnLeuGluLys-----LeuThrSer 190
Db 292200 TTATCGGAAAAAGGATGTTTCTGTATTTTCGGCAAAATGACAAATTCGATGTTCTTATAGT 292259
Qy 191 Gly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValIleHisThrGlu 206
Db 292260 GGTTCCTATAATAATGCGGATAATTACGCACCTGTAAGGCAACACAGCTAATAATACC 292319
Qy 207 GlyLeuTyArgLysSerGlyAspTyAlaValProArgTyArg-----Asn 222
Db 292320 GCC---TATAAACACGTTTGGGGGCTTAGCAAAATTCGGTTGGCAAAATTAATGATCGGAAC 292376
Qy 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
Db 292377 CGGTGGAAATTATCCCGCGAAACTCGTTTAA-----AACAAACACACCAAGCAATAATG 292434
Qy 243 GlyTrpArgLys-----ArgPheTyArgArgThrTySerSer 255
Db 292435 AGTGGAAAAACGAACCTTACTAATGACAAATACAGATCAAAATCAAAAGTTCACGAGAC 292494
Qy 256 ArgArgAspGlnTyArgLeuProAlaHisSerHisGluTyArgAsp----- 271
Db 292495 AAAAAGACGATCTTCTCTCTCTCAACACCAACCATCCATCAGAAAGACAGAGATTTT 292554

```

```

; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:

```

QY 272 -----CysHisAlaAspIlelle-----TrpGlnLysSerLeu 282
 Db 292555 ACTCTAAAGTGAACACACGTTAGGTAGTGTATTTAACTGATCAACAAATTCCTG 292614
 QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
 Db 292615 ATCAAAGCGCGTATTAAAC-TATTAT-----TTAAGGCA----- 292649
 QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHis 322
 Db 292650 GATAATCTTATCTA-----AATACGCATATCGCACTGTAT 292685
 QY 323 AsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLys 342
 Db 292686 AACAAATAAACTATTGAGAAGAACACGCGTAAAGTCACTGCTGTAAGATCAGACTAA 292745
 QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
 Db 292746 TTGACTACCGAGGT-----ATAATTTACGTAATTTCTCGGAATTATCT 292790
 QY 359 -----AspTyrHisAspGluLysAlaGlyAspAlaVal 370
 Db 292791 CACATTTCTTTGTTATGGGTGCAATATATCGAGATAAAATCCGTACCGAA----- 292844
 QY 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle 390
 Db 292845 -----CGAGGCACAAACGCTAGCGATGCGAAGTTTCGAGCGGACCCCTAT 292889
 QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSer 410
 Db 292890 AATCGAATCAACACTACAGCGGT---TATTTA----- 292922
 QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
 Db 292923 -----ATGCCCATATTCG 292937
 QY 431 PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgVal 450
 Db 292938 CTATTTGGG----- 292946
 QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
 Db 292947 GAAAAATGTAGTTTCGCAAGTGTAGTTATGACCACTACGATACCTCAAGTAACACC 293006
 QY 463 LeuLeuAspArgGluAsnTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
 Db 293007 GTAAATAACAAGGATAATCATTA----- 293030
 QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
 Db 293031 -----TCTCCGCGCACAAATTAACCTGGATAGTAC-----AATGGTTAGAT 293075
 QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 Db 293076 TTTACTGCCAAATATAATGAAGCTTTCGAGCACCATCTATGCAAGAGCGATTTGTAGT 293135
 QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
 Db 293136 GGTGTCACCTTGGGCGCAAACTCTAGCGGTAGATCATCAATAGATTTGTAGCAAAAT 293195
 QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
 Db 293196 CCAATTTGGCCCTGAAACAGCGAAAATAAAGAAATTAACGGAATCTACATTTTGTAT 293255
 QY 547 AlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
 Db 293256 AGTCTGTTTAAACAAGCGGATAAAATCAAAATTTGAAGCGACTTATTTCGTAATGATGTG 293315
 QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
 Db 293316 AAAGATTTTATTAACTTAAATAATTTAATGATGCAAAAGACAAAGTGCAGGTGCA 293375
 QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595

Db 293376 AATCCAAATACAAATGAGCATTTGTCGCAAAAATTTCCAGTATCAA----- 293423
 QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluLeuTyrPheLysProThrPro 615
 Db 293424 AACATAACTAATGCCCGTTTAAAGCGGTATTGAATTGCAAGCTCAATACCAA---ACAGAA 293480
 QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
 Db 293481 CGTTTAACGCTA-----TTTACTAATCTATGGCAGC 293510
 QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
 Db 293511 ACCAAAGTAAAGATAAAGATAGTGGC-----GAAGCT 293543
 QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
 Db 293544 TTATCAACATTCGCCAAGCAAAATCGCGTAAATGCTAGGGTAAATTTATGTTAGTAAAGAC 293603
 QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
 Db 293604 AATTCACGTTGGGAGCGACAGTAACCCATTACGCT-----GCTCAACGCCGATGCT 293657
 QY 693 ArgTyrGluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
 Db 293658 AAGATCATAGTGTACTACCCAGTATATATACTGACCGATCTTCGTGCTACCTATGCA 293717
 QY 712 ArgAsnThrArgTyrGlyGluTrp 719
 Db 293718 -----CCATTAAAGGCGAATGG 293735

RESULT 3

US-10-066-551-8
 ; Sequence 8, Application US/10066551
 ; Publication No. US20030100071A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Apicella, M. A.
 ; APPLICANT: Edwards, J. L.
 ; APPLICANT: Gibson, B. W.
 ; APPLICANT: Scheffler, K.
 ; APPLICANT: Brown, E.
 ; TITLE OF INVENTION: Vaccine and compositions for the
 ; prevention and treatment of Neisserial infections
 ; FILE REFERENCE: 875.045U1
 ; CURRENT APPLICATION NUMBER: US/10/066,551
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: US 60/344,452
 ; PRIOR FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: US 60/310,356
 ; PRIOR FILING DATE: 2001-08-06
 ; PRIOR APPLICATION NUMBER: US 60/266,070
 ; PRIOR FILING DATE: 2001-01-31
 ; NUMBER OF SEQ ID NOS: 12
 ; SEQ ID NO 8
 ; LENGTH: 2307
 ; TYPE: DNA
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-066-551-8

Alignment Scores:

Pred. No.:	2,44e-14	Length:	2307
Score:	227.50	Matches:	189
Percent Similarity:	33.14%	Conservative:	101
Best Local Similarity:	21.60%	Mismatches:	299
Query Match:	5.64%	Indels:	287
DB:	14	Gaps:	43

US-09-936-377-2 (1-758) x US-10-066-551-8 (1-2307)

QY 16 IleAsnThrProLeu-----LeuAla 22
 Db 99 ATGAATACCCCATTTGTCGCTCAGCCTGCTCGTCACTTGGCGAGGTTTGGC 158

Qy 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42
 Db 159 CACGGCGCAGAAATAATGCAAGGTGCTAGTACCGTACTGTGTAAGGCGACCGC 218
 Qy 43 ArgProArgAlaThrSerGlyLeuLeuHisThrThrAlaSerAspLysLileSer 62
 Db 219 CAA-----GGCAGCAAAATCCGTACCAACATCGTT--- 248
 Qy 63 GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp 78
 Db 249 -----ACGTGCAACAAAAGACGAAAGCAGCACCGATATGCGCACTTTAAA 302
 Qy 79 GlyValProGlyLileHisAlaSerGlnTrpGlyGlyAlaSerAlaProValile--- 97
 Db 303 GAAGAGCGTCCATC-----GATTTCGGCGCGCAACGCGACGTCCCAATTCCGT 353
 Qy 98 -----ArgGlyGlnThrGlyArgArgLileysVal----- 107
 Db 354 ACGTTCGGCGCATGGGTGAGAACTCTGTGACATCAAGGTGAGAACGCGCTATTCCGAC 413
 Qy 108 -----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 Db 414 AGCCAAATCCTTTTACCACCAAGGCAGA----- 440
 Qy 125 AlaIleMetValAspThrAlaLeuSerGlnValGluLileLeuArgGlyProValThr 144
 Db 441 ---TTTATTGTCGATCCGCTTTGGTTAAAGTCGTTTCCGTACAAAAGGC----- 488
 Qy 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLilePro 164
 Db 489 -----GCGGTTTCGCTCTCGCGGTATCGCGGACCAACGCGCGCATCATC 536
 Qy 165 GluLysMetProGlu-----AsnGlyValSerGlyGluLeuGlyLeuArg 179
 Db 537 GCCAAACCGTCGATGCCAAGACCTGCTCAAGGCTTGGATAAAATCGGGCGTGGC 596
 Qy 180 LeuSerSerdly-----AsnLeuGluLysLeuThrSerGlyGlyLileAsnLileGlyLeu 197
 Db 597 CTCAACAGCGGCTTTGCGACCAAGGCGGTAAAGTACGCGCAAGCGTATTTCGAAAA 656
 Qy 198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyr-----ArgLysSerGly 213
 Db 657 GAGGGCAACTTC-----GACGGCTTGTTCTTACACCGCAACGACGTAAGAAA 704
 Qy 214 AspTyrAlaValProArg-----TyrArgAsnLeu-----LysArgLeuProAspSer 229
 Db 705 GATTACGAAGCGCGCAAGAGTTTCCGCAATGTCAACGCGCGCAAAACCGTACCGAC 764
 Qy 230 Pro-----ArgargPheAlaAsnGlyGlnHis 238
 Db 765 GCGCTGGACAAACGAGTACTCTGCCAAATCGGAACACCTTCGCGAGCAGCAC 824
 Qy 239 ArgAlaValLeuGlyTyrArgLysArgPheTyrArgArg----- 251
 Db 825 CGCATGTTGTTGACCCACATGAAAGACCAACACCGGGCATCCGCACTGTGCGTGAAGAA 884
 Qy 252 ---ThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp 270
 Db 885 TTTACCGTCGCGCAAAAAGTTTACGGATAATATTGACCGCCCAAGCCCTGCTTACCGC 944
 Qy 271 AspCysHisAlaAspLileIleTyrGlnLysSerLeuLileAsnLysArgTyrLeuGlnLeu 290
 Db 945 GAA----- 947
 Qy 291 TyrProHisLeuLeuThrGluAspValAspTyrAspAsnProGlyLeuSerCysGly 310
 Db 948 -----ACTACCAATCCACACACCACTTGGGTATACGCGGTAAACCTCTGGC 995
 Qy 311 PheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyrPileAspLeu 330
 Db 996 TTTTCGAAAAAATCGGATGCCAAGCCTATGTGTT---GGAAAAAGAACGCTATTCCGC 1051
 Qy 331 ArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAla 350

Db 1052 CGATGACAGCGGACCGGCTTACGAGCAATGTAAAGGCCCAACCA----- 1099
 Qy 351 LeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaVal 370
 Db 1099 ----- 1099
 Qy 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgileGluLeuArgHisGlnPro-Il 390
 Db 1100 -----TACCCGAATCACCCTCGTGTGCGAACTTCAACTTCGACAGCGCCTTGC 1150
 Qy 390 eGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerAlaLeuSe 410
 Db 1151 CGAACAAACCTCTTGAATACGATATCACTACCGCCATCAGGAA----- 1196
 Qy 410 rAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSe 430
 Db 1197 -----ATCAACCGCAGCATTTTGAATCGAAA-----TTCTC 1231
 Qy 430 rPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgVa 450
 Db 1232 CATCCCGACGACAGAGAGAAAAAC-----GGTCAAAAAGT 1267
 Qy 450 lGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTy 470
 Db 1268 CGATAAACCGATGGAAACAACAATGAAGACCGTGCAGATCAAGACACTGTTCACGCTA 1327
 Qy 470 rLysGlnProLeuPro-----AspLeuGlyAlaHisArgGlnThrAlaArg----- 485
 Db 1328 CAACTTTCACCCGACCAACCGATACCGGTATATGTTGAAGCCATTACGACAT 1387
 Qy 486 ---SerPheAlaLeuSerGlyAsn----- 492
 Db 1388 CGGCGATTTCACGCTGACCGCGGCTGCTTACGACCGCTTCAAGGTGAAACCCATGA 1447
 Qy 493 -----TrpTyrPheTh 496
 Db 1448 CGGCAAAACCGTTTCAAGCAGCAACCTTAACCCGAGTTTCGGTGTGATTTGG-----CA 1501
 Qy 496 rProGlnHisLysLeuSerLeuThrAlaSerHisGln-----GluArgLeuProSerTh 514
 Db 1502 GCGCAGCAACACTGGAGCTTCAGCGGAGCCCAACTACGCCAGCGCGACCCCGCGCT 1561
 Qy 514 rGlnGlu---LeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPh 529
 Db 1562 GTATGACGCGCTGCAAAACCCACCGTAAACCGCGCATCATCTCGATTGCGACGCGACA-- 1619
 Qy 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
 Db 1620 -----AAAGCCGAACGCGCGCAATACCGAAATCGGCTTCAA 1657
 Qy 549 yTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
 Db 1658 CTACAACGACGCGCAGCTTTCGCCAAG-----GGCAGCTA 1693
 Qy 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGl 589
 Db 1694 CTTCTGG---CAGACCATCAAGACGCGCTTCCAAATCCGCAAAACCGCCACGACTCTGT 1750
 Qy 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyGl 608
 Db 1751 CGCGCTCGGTGAAGCCGTCATGCGGTATACATCAAAACCCAGCTTACGAAATGGGCGC 1810
 Qy 608 uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl 628
 Db 1811 GTCTACCGCAGCGCGCGCTGCTGCTCAAGTCCGCGTACG----- 1853
 Qy 628 yArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
 Db 1854 -----CACAGCAAAACCGCGCTTTACGATACGCAACAGACAGCTGTT 1897
 Qy 648 eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668

Db	961	GCCTGCGGCGGCAAAAT	---	---	TCA	981
Qy	281	SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVal	---	---	---	300
Db	982	CGGATAACTATTAAACGCCAAGCCCGCTGCTACCGGAA	---	---	---	1038
Qy	301	AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	---	---	---	320
Db	1039	AAATTGGCGTACACCGGCAAGATTGGGCTTTGTCGAAAACTGGATGCCAAGCCTAT	---	---	---	1098
Qy	321	AlaHisAsnGlyLysProTyrPheLeuAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu	---	---	---	340
Db	1099	GTGCT---GGAAAAAAGCGTATTCCGCGATGACAAAGATAACCGCTACGCGAGCAA	---	---	---	1154
Qy	341	TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr	---	---	---	360
Db	1155	TGTAAAGGCCCCNACCA	---	---	---	1193
Qy	361	HisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsn	---	---	---	380
Db	1194	CAT	---	---	---	1196
Qy	381	AlaArgIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValG1	---	---	---	400
Db	1197	---GAACCTTCAACTTCACAGCCGCTTGGCGAACAAACCCCTGTTGAATACGCGCATCAA	---	---	---	1253
Qy	400	nTyrLeuGlyGlnLys	---	---	---	407
Db	1254	CTACCGCATCAGGAATCAACCGCAAGCGTTTTGAATTCACAATTTAAATTTGAAGA	---	---	---	1313
Qy	407	rAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValG1	---	---	---	427
Db	1314	TAAAAAAGATGCAACTGAGGAAGATATAAAGAAAGAACCGTGAATAATGAAAAATTTGCCAA	---	---	---	1373
Qy	427	nHisTyrSerPheGlyValGluGlnAlaAsnTrp	---	---	---	439
Db	1374	AGCTACCGTCTGACCAACCGCAGCAAAACCGATACCGCGCGTATATCGAAGCCATTCA	---	---	---	1433
Qy	440	---AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerI1	---	---	---	457
Db	1434	CGAGATTGACGGCTTTACCTGACCGCGGCTGGTTCAGCAGCGCTTCAAGTGAAC	---	---	---	1493
Qy	457	eArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys	---	---	---	476
Db	1494	CCAC	---	---	---	1534
Qy	476	pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT	---	---	---	496
Db	1535	GTTT	---	---	---	1576
Qy	496	hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG	---	---	---	516
Db	1577	GCGGAGCCACAATAC	---	---	---	1621
Qy	516	luLeuTyrAlaHisGlyLysHis	---	---	---	532
Db	1622	CGTGTCAAAACCGGCAAAACCGGCAATCTCGATTGCGCGAGCC	---	---	---	1671
Qy	532	lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG	---	---	---	552
Db	1672	---AAGCCGAACCGCGCGCGCAATACCGAATCGGCTTCACTACACAG	---	---	---	1717
Qy	552	lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg	---	---	---	563
Db	1718	ACGGCACGTTTGGCGCAAAACGGCAGCTACTTCCGCGCAGACCATCAAGACGCGCTTGCCA	---	---	---	1777
Qy	564	---AspArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly	---	---	---	578
Db	1778	ATCCGCAAAACCGGCACGACTCTGTGCGCGTCCGCGAAGCGGTCAACCGCGCTACATCA	---	---	---	1837
Qy	578	---	---	---	---	578

RESULT 5

US-10-085-959-54
; Sequence 54, Application US/10085959
; Publication No. US20030165870A1
GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
; LENGTH: 32159
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10795)..
; OTHER INFORMATION: Unsure
US-10-085-959-54
Alignment Scores:
Pred. No.: 7.4e-12 Length: 32159
Score: 220.50 Matches: 187
Percent Similarity: 34.05% Conservative: 97
Best Local Similarity: 22.42% Mismatches: 309

Query Match:	5.46%	Indels:	242
DB:	12	Gaps:	45
US-09-936-377-2 (1-758) x US-10-085-959-54 (1-32159)			
QY	1	MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuAsnThrProLeu	20
Db	29138	ATGAATAAGACCGCTTATCCTCGGCTTTGGGGATTATTCGTC-----CCGCC	29191
QY	21	LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly	40
Db	29192	ATTGCTAATGCCAGACTTCACAGAAC-----GAAAGCACGTGTGTGTACC	29242
QY	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspIle	60
Db	29243	GCCAGTAACAATCTTCGGCTCGCA-----TCAGCCAACAACGCTCATCTACTGT	29296
QY	61	IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal	80
Db	29297	GTCCAGCGCGCGAAATTAAGCAGCGCGCGTCACCGCAGCGACAAATCCCGAGATC	29356
QY	81	---ProGlyIleHisAlaSerGlnTyrglyglyAlaSerAlaProVal---IleArg	98
Db	29357	TTCGCCGGCTCAATATTGAAATAGCGCAACATGCTTTTCGACGATCTCGTACGC	29416
QY	99	GlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMeta	118
Db	29417	GGCGTCTCTTCAGCGCAG-----	29434
QY	119	AspPhe---SerProAspHisAlaIleMerValasp-----	129
Db	29435	GACTTCTAATACCCCGCGCTACCGTGATGTCGATGGGTCTCAGCTTCCACCAAC	29494
QY	130	-----ThrAlaLeuSer-----GlnGlnValGluIleLeuArgGlyProValThrLeu	145
Db	29495	ACCATCCAGCGCTTACCGATGTCGAAAGCTGGAGTTGCTGGAGGCCACAGGAAACG	29554
QY	146	LeuTyrsSerSerglyAsnValAlaGlyLeuValAspValAlaAspGlyIlysileProGlu	165
Db	29555	TTATATGGCAAAAAGCGCTCAGGGCGGGATCATCAATCGTCAACCGAGCAG--CCGGAC	29611
QY	166	LysMetPro-----GluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSer	181
Db	29612	AGCAGCGCGCGCGCTATATTGAAGCGCGCTCAGTACGCGACATGTTATCGA---AGT	29668
QY	182	SerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPhe	201
Db	29669	AAGTTCAACCTG-----AGCGGCCCAATTCAGATGGCTG-----	29704
QY	202	ValLeuHisThrGluGlyLeuTyArgLys-----SerglyAspTyrrAlaValProArg	219
Db	29705	CTGTACGGCAGCGTCACCTGTTACGCCAGGTGTGATGACGGCAGCATGATTAA-CCCGCG	29763
QY	220	-----TyrArgAsnLeuLysArgLeuProAspSerProArg-----ArgPheAlaAsn	235
Db	29764	GACGGGAAGCGATGACTTAGCGGCACCCCGCCAGCATAGGGAATGTGAACCTGGCTCT	29823
QY	236	GlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyrrArgArgThrTyrrSerAsp	255
Db	29824	GGCGCCGACATCAGCCCTGGGA-----	29847
QY	256	ArgArgAspGlnTyrrGlyLeuProAlaHisSerHisGluTyrrAspCysHisAlaAsp	275
Db	29848	-----AATGGGCTT-----TGC-----	29859
QY	276	IleIleTrpGlnLysSerLeuIleAsnLysArgTyrrLeuGlnLeuTyrrProHisLeuLeu	295
Db	29860	-----CGCTCAGCGGAATGTACCGCGC-----	29883
QY	296	ThrGluGluAspValAspTyrrAspAsnProGlyLeuSerCys-GlyPheHisAspAspas	315
Db	29884	-----CACCAGAGATGCCATGTGGATGGGAATGATATTAA-----	29919

```
Db 30808 CAGTGAAGTTGATCAGCGTAACCGG-----GTGCCGTT 30840
Qy 658 gValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgileAspAl 678
Db 30841 CGTACACAGTTATGGCCGGGAACAGCGCGGTGATGATGATGATGATGATGATGATG 30900
Qy 678 aAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgTh 698
Db 30901 A-----CTGATGCCCGGACTGGCGGTTAATCT 30927
Qy 698 rProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyG 718
Db 30928 GGTGGCGCGCATATTTCGAT---GGCGAACACCGTTCGGCAAGGCACCTATGCCAC 30984
Qy 718 u-----Ttp-----AsnTrpTyrValLysAlaAs 726
Db 30985 CCTGGACAGACGCTGGGCTGGCAGGCGACTGAACGGATGACATTCCTCTATGCGA 31044
Qy 726 pAsnLeuAsnGln-----SerValTyrAlaHisSerSerPheLeuSerAspThrPr 744
Db 31045 TAACCTGTTCACCGTGGTACCGTACCTATGGCTACATGAACGCGCAGCGCGCTGCG 31104
Qy 744 oGlnMetGlyArgSerPheThrGlyGlyValAsnValLys 757
Db 31105 GCAGGTCAATATGGTGCACCGCTCGGTATCAATACGCGA 31144

RESULT 6
US-09-801-451A-1
; Sequence 1, Application US/09801451A
; Publication No. US20030096368A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule
; and Vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/801,451A
; FILING DATE: 08-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/418,964
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-3-P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 318..2456
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-801-451A-1
Alignment Scores: 5,16e-12 Length: 2600
Pred. No.: 207.00 Matches: 196
Score: 34.66% Conservative: 100
Percent Similarity: 22.95% Mismatches: 308
Best Local Similarity: 5.13% Indels: 254
Query Match: 11 Gaps: 41
DB: 41

US-09-936-377-2 (1-758) x US-09-801-451A-1 (1-2600)
Qy 11 LeuSerIleLeuLeuIleAsnThrProLeu---LeuAlaGlnAlaHisGluThrGluGln 29
Db 339 CTCAGCTGCTCTCGCTCACACTTTCGCGCGCGCTTTCGCCACGCGGAGAAAATAATGCC 398
Qy 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49
Db 399 AATGTCGATTTGGATACCGTTACCGTAAAGAGGAGCCGCCAA----- 440
Qy 50 LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLys 69
Db 441 -----GGCAGCAAAATCCGTACCAACATCGTT-----ACGTTCAACAAAA 482
Qy 70 -----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
Db 483 GACGAAAGACCGCAACCGATATCGCGAACTCTTAAAGAGAGAGCCCTCCATC----- 536
Qy 86 SerGlnTyrGlyGlyGlyAlaSerAlaProValIle-----ArgGlyGln 100
Db 537 ---GATTTTCGCGCGCGCAACGCGACGTCCTCCCAATCTCTGACGCTGCGCGGTATGGTCA 593
Qy 101 ThrGlyArgArgIleLysVal-----LeuAsnHisHis 111
Db 594 AACTCTGTCGACATCAAGGTGGACACACGCTATTTCGACAGCAAAATCTTTACCACCAA 653
Qy 112 GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla 131
Db 654 GGCAGA-----TTTATTGTCGATCCCGCT 677
Qy 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db 678 TTGTTAAAGTCGTTTCGTACAGAAAGC-----GCGGTTTC 716
Qy 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu----- 169
Db 717 GCCTCTGCGGTATCGCGCGACCAACGCGCGCATCATCGCAAAACCGTCGATGCCCAA 776
Qy 170 -----AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSer----- 182
Db 777 GACCTGCTCAAAAGGCTTGGATAAAACCTGGGGCGTGGCGCTCAACAGCGGCTTTCGCGGC 836
Qy 183 -----GlyAsnLeuLysLeu 188
Db 837 AACACGCGCAAGCTACGCGCAAGCTATTTCGAAAAGAGGCACTTCGCGGTTTG 896
Qy 189 ThrSerGlyGlyIleAsn-----IleGlyLeuGlyLysAsnPhe-ValLeuHi 204
Db 897 TTCTCTTACAACCGCAACGATGAAAGATACGAAAGCGCGCAAGGTTTCGCAATGAC 956
Qy 204 sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeu 224
Db 957 AACG-----GCGGCAAAACCGCTACCGCTGACGCGGCTCGCAAAACGACG 1001
Qy 224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
Db 1002 TACTTCGCCAAAATCGGAACAACCTTCGCGCGAGCGGCAACCCATCTGTTGAGCCA 1060
Qy 244 pArgLysArgPheTyrArg-----ArgThrTyrSerAspArgAspGlnTyrGlyLe 262
Db 1061 TATGAAAGACCAACACCGCGGCGCATCCGCACTGTG-----CGTGAAGAGTTTGC 1111
```


Db	1962	GC	GTCAACGCGCGGTACATCAAAAACCGGTTACGAATTTGGCGCGGTCTCTACGCACC	2021
Qy	613	ProThrProArgTyrArgIleGlyValSer	-----	622
Db	2022	GGCGGCTGACCGCAAGTCGGGTAAGCCACAGCAACCGCGCTTTTACGATACGCAC	2081	
Qy	623	-----	-----GlyAspTyrValArgGlyArgLeuLys	631
Db	2082	AAAGACAAGCTGTTAGCGCGAACCCCTGGAATTTGGCGCACAAAGT-CGGCGGCACCTTGGAC	2140	
Qy	632	AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIlealaGln	651	
Db	2141	GGCGCTCCCTTGCCTACCGCTTCAAAAAACCCGAAATCTGGAAATCGG-----CTGCGCGCGG	2194	
Qy	652	AspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHis-LeuLysAlaLe	671	
Db	2195	TCGTTATGTTCAAAAACCGT-----	-----GGGTTCGATATTGGCGGCAGG	2236
Qy	671	rLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLe	691	
Db	2237	TCAAAAAGACCGC--GACGCGAAATTTGGAA-----	-----AAGCTTGTACGCCAA	2279
Qy	691	uAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrAr	711	
Db	2280	-----	-----GGTTCCGGTGTGAACGATGCTTCGCCCAACTCGAA	2314
Qy	711	g-----	-----ArgAsnThrArgTyrGlyGluTTPAsnTyrValLysAlaAspAsnLe	728
Db	2315	ACCGCTGGCAAGACAGC-----	-----CTCAATGTTAATCTTTCCGGTTAACCAACGT	2362
Qy	728	uLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGl	747	
Db	2363	GTTCGACAAAGTCTACTATCCGACACGCCAACGCTGGACCAATACCTGCGCGGCGTGGG	2422	
Qy	747	YArgSerPheThrGlyGlyValAsnValLysPhe	758	
Db	2423	ACGTGATGTACGCTGGCGCGTGAACTACAAGTTC	2456	
RESULT 7				
US-10-181-319-1				
; Sequence 1, Application US/10181319				
; Publication No. US20030195032A1				
; GENERAL INFORMATION:				
; APPLICANT: Lewis, Thomas A.				
; APPLICANT: Paszczynski, Andrzej				
; APPLICANT: Crawford, Ronald L.				
; APPLICANT: Cortese, Marc S.				
; APPLICANT: Sebat, Jonathan L.				
; TITLE OF INVENTION: Compositions and Methods for Bioremediation				
; FILE REFERENCE: IDAHI19440				
; CURRENT APPLICATION NUMBER: US/10/181,319				
; CURRENT FILING DATE: 2002-07-15				
; PRIOR APPLICATION NUMBER: PCT/US01/02386				
; PRIOR FILING DATE: 2001-01-19				
; PRIOR APPLICATION NUMBER: 60/177,251				
; PRIOR FILING DATE: 2000-01-20				
; NUMBER OF SEQ ID NOS: 35				
; SOFTWARE: PatentIn version 3.0				
; SEQ ID NO 1				
; LENGTH: 2061				
; TYPE: DNA				
; ORGANISM: Pseudomonas stutzeri				
; FEATURE:				
; NAME/KEY: CDS				
; LOCATION: (1)..(2061)				
; OTHER INFORMATION: ORF K				
US-10-181-319-1				
Alignment Scores:				
Pred. No.: 3,11e-11 Length: 2061				
Score: 198.50 Matches: 161				
Percent Similarity: 36.08% Conservative: 106				

Alignment Scores:	
Pred. No.:	3.11e-11
Score:	198.50
Length:	2061
Matches:	161
Percent Similarity:	36.08%
Conservative:	106

Best Local Similarity: 21.76% Mismatches: 283			
Query Match: 4.92% Indels: 192			
DB: 12 Gaps: 36			
US-09-936-377-2 (1-758) x US-10-181-319-1 (1-2061)			
Qy	42	SerArgProAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle	61
Db	179	TCACAGCCGAGAAATCGAGCGCG-CTGAAAGGGTGCCTCCACCGTGGCGGTATC	237
Qy	62	SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly--Val	80
Db	238	GATGGCTGGACCGCGAGCAGTCAGGCATCACTAGCCTCAACAACTGGAAGCAGCAT	297
Qy	81	ProGlyIleHisAlaSerGlnTyrGly---GlyGlyAlaSerAlaProValIleArgGly	99
Db	298	CCTGGTCTCTCATTCACAGCGTTCGGGCAAGCAGGTATGAATTCACCGCTCATCGGGG	357
Qy	100	GlnThrGlyArgArgIleLysValLeuAsnHisHISgLyGlnThrGlyAspMetAlaAsp	119
Db	358	CTG-----ACGGCCAACTTCAACAGC	378
Qy	120	PheserProAspHisAlaIleMetValAsp-----ThrAlaLeuSerGln-----	134
Db	379	TTCTCCAGTTCAACGTTGTTGCTGTCGATGGCGTTCACGCTGACAGCCAGGGATTC	438
Qy	135	-----GlnValGluLeuLeuArgGlyProValThrLeuLeu	146
Db	439	GAGAGTGGCATGTGGATTCGATCGCATCGAGTTCATTCGGCGCGCAATTCATCGCTG	498
Qy	147	TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro---Glu	165
Db	499	TATGGCGGTAAATCCGAGGCGGGTGTG-----ATTGCCATCCACGCTGCCGATGAC	552
Qy	166	LysMetProGluAsnGlyValSerGlyGluLeuLeuArg-----	179
Db	553	CGACACCGGAGACGCGTGTCTGCGCAAGCGCGGACCGGAACAACGCGTGTATCGCG	612
Qy	180	-----LeuSerSerGlyAsnLeuGlnLysLeuThrSerGlyGlyIle-----AsnIle	195
Db	613	TTTGGCGTCCAGCCAGCCTTGGTGGAAAGCGGTTGTACGCGACGCTATCGGCAACTGG	672
Qy	196	GlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyr	215
Db	673	TCGAGCCAGGACGCGTTTCATCGACACACCCAC---ACGGGGCACAGGCGGACGAT---	726
Qy	216	AlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsn	235
Db	727	-----CGTGACGAGAAACCTG---AACCTGGGGCTGCGTGGCGCCCG	768
Qy	236	GlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAsp	255
Db	769	GGGGCCCAACGATGTGTC-----ATGGCTATGGCATCAGGATGACGAGAT	819
Qy	256	ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAsp	275
Db	820	GGCGCCTCCCTGTGGGCTCGCCGCGCGCAAGGAAG---CGAGTCGCGTCCGGAAG	876
Qy	276	IleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu	295
Db	877	CCGAGCTGGAAACCGTTCT-----	894
Qy	296	ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAsp	315
Db	895	-----GAGGGCAGACCTTGTCTTCAATGTCCAGCATGAATTT	933
Qy	316	AspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAsnLeuArgAsnLysArgTyr	335
Db	934	GCCTCCGCGCTCGCGTTGTCATTCGGTAACGCGCTGGAACGAGTTTCAAGGAC-----	984
Qy	336	GluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu	355
Db	985	-----AGGATTACAGAGGACACTGACTTATGCCAGCCGATGTCTG-----CACGTC	1032
Qy	356	AsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsn	375
Db	1033	GGGCGCGAC-----CATCAC-----CTGCGC	1053
Qy	376	AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeu---Lys	394
Db	1054	ACACTCTCCAGGAGTTCGTTGGAG-----GGACAGCTCGGGAG	1095
Qy	395	GlySerTyr-----GlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThr	412
Db	1096	GCAGTTTGGCTGGCTGGTGC---TAGCGGATCGGACGACACAGTCTGCACAGTACC	1152
Qy	413	SerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePhe	432
Db	1153	AGCAAGACCATGATGGG-----CTGTCGACATTCGCGCGGATCAGCAGCGATACC	1206
Qy	433	GlyValGluGlnAlaAsnTyrAsp-----AsnPheThrLeuGluGlyGly	447
Db	1207	GCTGCACCTGTCCACCACTGGAACGTCCTCCCTGTCGCCGACCTGGTCCATAGCCCGGA	1266
Qy	448	ValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGlu	467
Db	1267	CGCGCGTTCGAGCGCAAGGAGTGCAGTACGT---CCGCAAGGGGTACGAGCCATGAA	1323
Qy	468	AsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPhe	487
Db	1324	AAAGGCTGGACACACGCTTTCACCCAGGCTCGCGTGCAACACACAG-----	1368
Qy	488	AlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis	507
Db	1369	-----ATAACGCCCAATCAC	1383
Qy	508	GlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn	527
Db	1384	CAATGGTATGTAGTCCAGTCCGTCGTCGCGCTCGGCTTCATGTGCTGGCGCG	1443
Qy	528	ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAlleGluLeuAla	547
Db	1444	ACGCTG-----GGTTATCTGCTTACGACGAGGAGAACTGGTCTGATGAA	1491
Qy	548	LeuGlyTyrGluGly-----AspArgTyrGlnTyrAsnLeuAlaLeuTyrArg	563
Db	1492	ACCGTCTCAGGCGCTGGCTTTCGACAGCGATTCGCTATTTCGTCGCCGCTAC---	1548
Qy	564	AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer	583
Db	1549	-----CTCATGGAC-----	1557
Qy	584	IleGluAspAspSerGluMetLysLeu-----ValArgTyrAsnGlnSerGly	599
Db	1558	ATCGATGACATGCAAGTTCATGCGAGATGCCACCGTCGCGCTGATGATCATCCAGCGCT	1617
Qy	600	AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle	619
Db	1618	GCCACGCGCATCCAAAGTCTCGAGCTGGATGGACTA-TCTCTCGGTGGCGGCTG	1676
Qy	620	Gly-ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyAr	639
Db	1677	GCAGTCAAGGCGGGCTGGCTGGAAACACACCGCTTCGATCATCTTCGATGGCGA	1736
Qy	639	GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAsp-----	652
Db	1737	GGCGGACTATGACGGCAACCAAGCCGTTTCGCGCGGATCTCATCCGCGCCACCTCGGCAT	1796
Qy	653	-AspGlnAspArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLe	672
Db	1797	CCGCTAGCAGCCCGCGAA-----GGCTGGTATGCAACAGCCGCT	1838
Qy	672	uThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAl	692
Db	1839	GACCGCGAGCAGCAAGTCTACCTCGATGCGGCGCAACGGGTATGAAACGCAAC-----	1890

QY 692 aArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAr 712
Db 1891 -----GGCTACGGCGCTGGGAACTGGTAGCTGGTTACCAAG 1928
QY 712 gAsnThrArgTyrGlyGluTrpAsnTrpTyrValIysAlaAspAsnLeuLeuAsnGln 731
Db 1929 C-----GGCAACTGGGAATCGGGCTACGCCGCAACACGCGACCGATCAG 1974

RESULT 8

US-10-114-170-194/c
; Sequence 194, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9057
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-10-114-170-194
Alignment Scores:
Pred. No.: 4,05e-09 Length: 9057
Score: 188.00 Matches: 162
Percent Similarity: 33.47% Conservative: 258
Best Local Similarity: 21.43% Mismatches: 248
Query Match: 4.66% Indels: 248
DB: 14 Gaps: 33

US-09-936-377-2 (1-758) x US-10-114-170-194 (1-9057)
QY 4 ThrThrLeuLysProIleValLeuSerIleLeuLeuLeuAsnThr---ProLeuLeuAla 22
Db 3100 ACCTGTTGCGTTTGGATTATGGCTTAGCTGTTCTGCCACCTTGCCACCGTTTGCT 3041
QY 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42

Db 3040 TTTGTACTGAAACC-----ATGACCCTTACGGCAACGGGAATGCC 2999
QY 43 Arg-----ProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSer 57
Db 2998 CGTAGTTCTTCGAAGCGCTATGATGGTCAGC---GTCAATCGACACTTCGGCTCCTGAA 2942
QY 58 AspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu 77
Db 2941 AAT-----CAAACGGCTACTTCAGCCACCGATCTGCTG 2909
QY 78 AspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal--- 96
Db 2908 CGTCATGTTCTCGGAATTACTCTGAGTACGACGACCAACACGGTCAGGATGAAAT 2849
QY 97 IleArgGlyGlnThrGlyArgArgIleLysValLeu-----AsnHisHisGly 112
Db 2848 ATGCGGGCTATGATCATCGCGCGTGGTCTTGTGATGGTGTTCGTACGGGAACG 2789
QY 113 GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
Db 2788 GATACCGGACACCTG-----AATGGCACTTTTCTCGATCCGGGCTG 2747
QY 133 SerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnVal 152
Db 2746 ATCAACGCGTTGAGATTGTTCTGGACCTTACGATTACTGATGGAGTGGCGGCTG 2687
QY 153 AlaGlyLeuVal-----AspValAlaAspGlyLysIleProGlyLysMetProGluAsn 170
Db 2686 GGTGGAGTGTCTCTACGATACGATCGGTCGATGCAAAA-----GATTATTTCAGGAA 2636
QY 171 GlyValSerGlyGluLeuGlyLeuArgLeu---SerSerGlyAsnLeuGluLysLeuThr 189
Db 2635 GGACAAAGCAGT-----GGTTTCTGTGTCTTGTGTACGCGGC----- 2597
QY 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu--- 208
Db 2596 ACGGGGACCATAGCCTGGGATTAGCGGAGCGCGCTTTGGCGGAACCTGAAATCTGGAT 2537
QY 209 -----TyrArgLysSerGlyAspTyrAla 216
Db 2536 GGTATTGTGCGCTGGTCCAGTCGCGATCGGGGTGATTATCGCCAGAGCAATGCTGAAACC 2477
QY 217 ValProArg----- 219
Db 2476 GCGCGGAATGACGAGTCCATTAAATACATGCTGCGAAAGGACCTGGCAATTGATTCA 2417
QY 220 -----TyrArgAsnLeuLysArgLeuProAspSer 229
Db 2416 GCCCAGTCTCTGAGCGGTTTAGTGGCTTACTACAACACGCGCGTGAACCAAAAAT 2357
QY 230 ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyr 249
Db 2356 CCGCAGACCGTTGGGGCTTCTGAAGCAGCACCCGATGGTGTGATCGTTCAACAATTCAA 2297
QY 250 ArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyr 269
Db 2296 CGC-----GATCGCAGCTTTCTTATAAATCGCCCGCAGCGACGACTGG 2249
QY 270 AspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGln 289
Db 2248 TTAATGTCAGATGCAAAAATTATTGTCGGAAGTCCGCTGATTAAATCGCAA----- 2198
QY 290 LeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys 309
Db 2197 -----AACACGGGGAGTTCGGC 2180
QY 310 GlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAsp 329
Db 2179 GAGTATCGTGAACA-GATAACAAAAAGGACGAGGCT-----GGAGAACCGTTCCACTCT 2127
QY 330 Leu---ArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPhe 348

523	Db	ACGGGGACCATAGCCTGGGATTAGGCGGAGCGGTTGGCGCAACTGAAATCTGGAT	582
209	QY	-----TyrArgLysSerGlyAspTyrAla	216
583	Db	GGTATTGTGCCTGCTCCAGTCGCGATCGGGGTGATTACGCCAGAGCAATGTTGAACC	642
217	QY	ValProArg-----	219
643	Db	GCGCCGAATGACGAGTCCATTAAATAACATGCTGGCGAAAGGAGCTGGCAATTTGATTCA	702
220	QY	-----TyrArgAsnLeuLysArgLeuProAspSer	229
703	Db	GCCAGTCTCTGAGCGGTTTAGTGGTACTACAAACACGCGCGGTGAACCAAAAT	762
230	QY	ProArgArgPheAlaAsnGlyGlnHisArgAlaValIleuGlyTrpArgLysArgPheTyr	249
763	Db	CCGACAGCCGTTGAAGCTTCGTATAGCAGCAACCCGATGTCGTCTGTTCAACAATTCAA	822
250	QY	ArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyr	269
823	Db	-----GATGGCGAGCTTCTTATAAATCGCCCGCAGGAGTAACGACTGG	870
270	QY	AspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGln	289
871	Db	TTAAATCCAGATGCAAAAATTTACTGTCGAGAGTCGTATTATATGCGCAA-----	921
290	QY	LeuTyrProHisLeuLeuThrGluuAspValAspTyrAspAsnProGlyLeuSerCys	309
922	Db	-----AACACGGGAGTTCAGGC	939
310	QY	GlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAsp	329
940	Db	GAGTATCGTGAACA-GATAACAAAGAGCAAGGCT-----GGAGAACCGTTCACACTCT	992
330	QY	Leu--ArgAsnLysArgTyrGluLeuArgAlaGlnTrpLysGlnProPheProGlyPhe	348
993	Db	ATTTCGCGACAGTTTCGCTCTCACTTACTGACAAATG-----	1031
349	QY	GluAlaLeuArgValHisLeuAsnArg-AsnAspTyrHisHisAspGluLysAlaGlyAs	368
1032	Db	-----CGTGAGTATTATCGTCAGGAACAACATCGGGTGG	1067
368	QY	pAlaValGluAsnPhe-----PheAsnAsnGlnThrGlnAsnAlaAr	382
1068	Db	CGCGACACGGGCTTCGCCAAGCAAAATCGATTTTAGCTCTGGTTGGCTACAGATGA	1127
382	QY	GileGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLe	402
1128	Db	GATCACCTTACGCGATCTGCCGATT-----	1152
402	QY	uGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLe	422
1152	Db	-----	1152
422	QY	uAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPh	442
1152	Db	-----	1152
442	QY	eThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAl	462
1153	Db	-ACCTGCTTGGGGAAACCGCTATGACATTTATCGGGTAGC-----	1194
462	QY	aLeuLeuAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgG1	482
1195	Db	-----AGCGACGGCTACAAA-----GATGTTGATGCCACCAATG	1229
482	QY	nThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSe	502
1230	Db	GTCACTCTCGTCGG-----GGGATGACTATCAACCGACCACTGGCTGAT	1274
502	QY	rLeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHi	520

Db 1275 GTTATTGGCTCATATGCTCAGGCATTTCGCGGCCCGCCGACGATGGCGGAATGTATTAACGA 1334

Qy 520 sGlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsnLysHisLe 536

Db 1335 TTCTAAACACTCTTCGATTGGTCGCTTCTATACCACTATTGGGTGCCAAACCCGAATT 1394

Qy 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly----- 549

Db 1395 ACGTCCGGAACTAAACAAACTCAGGAGTACGGTTTTGGCGTCGCTTTTGTGACCTGAT 1454

Qy 550 -TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569

Db 1455 GTTGTCCAATGATGCTCTCGAATTTAAAGCCGACTTTTGATACCAAGCGAAAGATTA 1514

Qy 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGl 589

Db 1515 TATC-----TCCACGACCGTCGATTTCCGGCG 1541

Qy 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla-----GluGl 607

Db 1542 GCGCACAACTATGTCGTATAACGTCGCCGAACGCAAAATCTGGGGCTG-GGATGTGATGA 1600

Qy 607 yGluIleTyr-----PheLysProThrProArgTyrAr 618

Db 1601 CGAAATATACCATGATCTGTTAGCCTTGATGTGGCTATAACCGTACCCTCGCGGCAAG 1660

Qy 618 gIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGl 638

Db 1661 ACACCGATACCGGGATATAT-----CTCCAGCATTAACCCCG 1699

Qy 638 yArgGluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaPro-A 658

Db 1700 ATACCGTTACCAAGTAC-----CCTGA 1720

Qy 658 rgValProAlaAlaArgLeuGlyPheHisLysAlaSerLeuThrAspArgIleAspA 678

Db 1721 ATATTCGATCGCTCACAGCGGCTTC----- 1746

Qy 678 laAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgT 698

Db 1747 --TCTGTTGGTTGGTCGGTACGTTGCCATCGCTCA----- 1782

Qy 698 hrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyG 718

Db 1783 -----ACACATATCAGCAGCAGCTACAGCAACAACTGGCTATGGTG 1825

Qy 718 luTrpAsnTrpTyrVal 723

Db 1826 TGAATGATTCTACGTC 1842

RESULT 10

US-10-085-959-172/c

; Sequence 172, Application US/10085959

; Publication No. US20030165870A1

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; APPLICANT: Burchland, Valerie A.

; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073

; FILE REFERENCE: 960296.97648

; CURRENT APPLICATION NUMBER: US/10/085,959

; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: 60/242,412

; PRIOR FILING DATE: 2000-10-19

; NUMBER OF SEQ ID NOS: 255

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 172

; LENGTH: 9054

; TYPE: DNA

; ORGANISM: Escherichia coli

US-10-085-959-172

Alignment Scores:

Alignment Scores:


```

Db      1515  CAAAATATACCACTGATCTGTTTAGCCTTGATGGGCTATACCGTACCCGCGCAAG 1456
Qy      618   gileGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGl 638
Db      1455  ACACCGATACCGGGGAATATAT-----CTCCAGCATTAACCCGG 1417
Qy      638   yArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaPro-A 658
Db      1416  ATACCGTTACCACTAC-----CCTGA 1396
Qy      658   rgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspA 678
Db      1395  ATATCCGATCGCTCACAGCGGCTTC----- 1370
Qy      678   laAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgT 698
Db      1369  --TCTGTTGGTGGTTCGCTACGTTGCCGATCCTCA----- 1334
Qy      698   hrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyG 718
Db      1333  -----ACATATACGACGACTACAGCAACACACCTGGCTATGGTG 1291
Qy      718   luTrpAsnTrpTyrVal 723
Db      1290  TGAATGATTCTACGTC 1274

RESULT 11
US-10-238-075-507/c
; Sequence 507, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 507
; LENGTH: 12264
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-507

Alignment Scores:
Pred. No.: 1.75e-08 Length: 12264
Score: 184.00 Matches: 134
Percent Similarity: 37.06% Conservative: 108
Best Local Similarity: 20.52% Mismatches: 236
Query Match: 4.56% Indels: 176
DB: 12 Gaps: 35

US-09-936-377-2 (1-758) x US-10-238-075-507 (1-12264)
Qy      1   MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db      10199  ATGTATATGAATGTAATCAGAACTGCAATTGTACATTAATATACTT-----CCGGTG 10146
Qy      21   LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
Db      10145  GGATTACAGCGCAGCCAGTCATTCTTCTATGTTAAAGATACAAATCACCATTGTTCGG 10086
Qy      41   LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db      10085  -----ACAGGAAATCAGAACACGCGTATTTGAAACGCGCTGATGGTCAGTGTC 10038
Qy      61   IleSerGlyAspThr---LeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly 79
Db      10037  GTCAGAAATGACACACCGTGGAGTCAAGATCGGTTACATCGGCCGCGCATGCTGAAGGT 9978

```


QY	411	a	ThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer	Ph 431
Db	9135	TGCACA	-----CTATATCCGAAGAGAAACATCTACATTTACATCGTT	9094
QY	431	e	PheGlyValIcLglnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgVal	1451
Db	9093	GTATTTCCAGGATGAATAGACAATGAAAGAGCTACCCGGTTAACATTATCGTC	-----TACAA	9019
QY	451	u	LysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr	471
Db	9041	-----GGTTTCCCGCTATGACCGG-	-----TACAA	9019
QY	471	s	GlnProLeuPro-----AspLeuGlyAlaHisArgGlnThrAlaArgSerPhe	488
Db	9018	GAGCTTCAATCCCGCGTCCGAGAGACTGAAGACCGCAACGCTGTCCTCCCAAGGCG--GC	8962	
QY	488	a	LeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu-----ThrAla	506
Db	8961	GAITTTCA-----GTCTCACCGACAGACTGGCTGATGATGACGGCTCCATATC	8914	
QY	506	r	HisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAla	526
Db	8913	CTCTGCAATTCCGACGCGCCACAAATGTCAGCAAAATACAGGAGATGATGTACATTTT	8859	
QY	526	r	AsnThrPheGluValGlyAsn-----LysHisLeuAsnLysGluArg	541
Db	8958	-----TACGCAAGGTTAAACCACTACTGGTTCCTAACCTTAATCTGAACACAGA	8806	
QY	541	r	AsnAsnIleGluLeuAlaLeuGly-----TyrGlu	552
Db	8805	AAATAACATCACCCGTCGAGATTGGCGCAGGTATTCAACTGGATGGCTGTCTTACAGACAA	8746	
QY	552	y	AspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr	572
Db	8745	TGACCGGCTCGAGTTAAAGCGCGATATTTCGGAACGAGTCCAGAAACTATATTGCCAC	8686	
QY	572	a	GlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGluMetLys	592
Db	8685	ACGGCTG-----GATATGAACGGATGCGCTTC	8659	
QY	592	u	ValArgTyrAsnGlnSerGlyAlaAspPheTyrGly	604
Db	8658	TTATTTCTATAATGATATCCGGGCCCGCTATCTGGGGA	8622	

RESULT 12
US-10-085-959-11
; Sequence 11, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 23654
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION: Unsure
US-10-085-959-11

[illegible]


```
Db 613 GGTAATATCTATCAAAAGTGATGGTGAGCACCGCA---CCTAAC-----AAGGAA 657
Qy 226 LeuProAspSerProArgArgPheAlaAsnGly-----GlnHisArg 239
Db 658 AAACCTGAGCC-----CTGTTTGGAAAGGCTCTGTCGGTATATAACAGACAGTAACAAA 711
Qy 240 AlavalLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln 259
Db 712 GCA-----GGTGCACGCTTGCCTCTCTACCGAATAACACCACTGAA----- 753
Qy 260 TyrGlyLeuProAlaHisSerHisGlnTyrAspCysHisAlaAspIleIleTyrGln 279
Db 754 -----CGGGCAATTCACCTAG-----ACACATGTGTAC----- 783
Qy 280 LysSerLeuLeuAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
Db 784 AGCGGCTGCTGCAGACAAAAACAGTACAA-----AATGAC 819
Qy 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAspAlaHisAla 319
Db 820 GTACAGTTCTG-GTACAGATGACGTCCTCTCTACCGAATAACACCACTGAA----- 851
Qy 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db 851 ----- 851
Qy 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db 852 -----TAACAGCCT-----CATCAATGTTAAAGTCAACGTTTATATCATGTA 893
Qy 360 TyrHisHisAspGluLysAlaGly-AspAlaValGluAsnPhePheAsnAsnGlnThrG1 379
Db 894 TATCACTATCAACAACCGTCTCAACAACCGCAGATGGAGAACACACAGAACCTC 953
Qy 379 nAsnAlaArgIleGluLeuArgHisGlnProIle-----GlyArgLeuLysGlyse 396
Db 954 CGGTGTTAATGTTCAACAGAGTACTACTCTGATTTTCCGGAGGCCCATCAGTTAAG 1013
Qy 396 rTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVa 416
Db 1014 TTATGGCGCTCAATATTACCGTCAGCAGCAGAACGACGAGCTCTGCCACA----- 1065
Qy 416 lLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluG1 436
Db 1066 -----CTATATCGGAAGAAACATTGACTTTACATCTGTTGTTATTCAGGAGGA 1115
Qy 436 nAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSe 456
Db 1116 AATGACAATGAAAGCTACCGCGTTTAAACATTATCGTC-----GG 1154
Qy 456 rIleArgTyrAspLysAlaLeuIleAspArgLysGlnTyrTyrLysGlnProLeuPro-- 475
Db 1155 TTCCCGCTATACCGG-----TACAAGAGCTTCAATCCCGC 1190
Qy 476 -----AspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTr 493
Db 1191 TGCCGGAGAACTGAAGCGGACCGCTCTCCCAAGGCG---GGATTCA----- 1239
Qy 493 pTy-PheThrProGlnHisLysLeuSerLeu-----ThrAlaSerHisGlnGluArgLe 511
Db 1240 ---GTCTCACCGACAGCTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295
Qy 511 uproSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVa 531
Db 1296 CCCACAAATGCCAGAAATGTACAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1343
Qy 531 lGlyAsn-----LysHisLeuAsnLysGluArgSerAsnAsnIleGluLe 546
Db 1344 GGGTAAACCAATTAATGCTTCTTAACTGAAACCAACCAACCAACCAACCAACCAACCAAC 1403
Qy 546 uAlaLeuGly-----TyrGluGlyAspArgTrpGlnTy 557
Db 1404 TGAGATTGGCGCAGGTATTCAACTGGATGGCTGCTTACAGACAATGACCGCTGCAGTT 1463
```

```
Qy 557 rAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAs 577
Db 1464 AAAAGGGCGGATATTTCGGAACGGATGCGAGAAACTATATTGCCACACGCGTG----- 1515
Qy 577 pGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnG1 597
Db 1516 -----GATATGAACGGATGCGTTCTTATTCTTTATAATGT 1550
Qy 597 nSerGlyAlaAspPheTyrGly 604
Db 1551 ATCCCGGCGCGGTATCTGGGGA 1572

RESULT 14
US-10-043-344-4
; Sequence 4, Application US/10043344
; Publication No. US20030088086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murdin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 5099
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (160)..(2121)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2152)..(4890)
; US-10-043-344-4

Alignment Scores:
Pred. No.: 1,14e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 14 Gaps: 48

US-09-936-377-2 (1-758) x US-10-043-344-4 (1-5099)
Qy 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuLeuAsnThrProLeu 20
Db 2155 ACTAAAAAACCCCTATTTTCGCTAAGTATTATTCTTTCTTTAAATTCATGCTATGTA 2214
Qy 21 LeuAlaGlnAlaHis-----GluThrGlnSerValGly----- 32
Db 2215 AAACGAGAACTCAAAGTATAAAGATACAAAGAGCTATATCATCTGAAAGTGGACACT 2274
Qy 33 -----LeuGluThrValThrValValGly-----LysSerArgpro 44
Db 2275 CAAAGTACAGAGATTTCAGAAATAGAACTATCTCAGTCTACGAGAAAATAAGAGAT 2334
Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIlelle-----Ser 62
Db 2335 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATCAAAACTAGT 2382
Qy 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
```

Db 2383 GAAAGTATCAGCCGAGAACAGTATTAAATTCGTGATCTAACCGCTATGATCCAGGC 2442

QY 83 lIeHiSaIaSerGlnTyrGlyGlyAlaSerAlaProVal---lIeArgGlyGlnThr 101

Db 2443 ATTTCAGTTGTAGAACAGCCGCTGGTGAAGTTCTGGATATCTTATTCGTGTATGGAC 2502

QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121

Db 2503 AGAAATAGAGTT----- 2514

QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGly 141

Db 2515 -----GCTTTATAGTAGATGGTTTACCTCAACGCAATCTTATGTAGTGCAAGC 2565

QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159

Db 2566 CTTTTAGTTGCTGCTGTCAGGATATCTGGCAATTCAGGCAATTAATGAATGAATGAA 2625

QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179

Db 2626 AATGTAAAGCCGTCGAA---ATAAGCAAGGGGGGAGTTCTTCTGAGTATGTTATGGA 2682

QY 180 LeuSerSerGlyAsnLeuLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198

Db 2683 GCACTAGCTGCTGTTCTGTAAATTCAAAGCAATTCAGCGCGATATCTTAGAAGGAGAC 2742

QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210

Db 2743 AAATCATGGGAATTCAAACTAAATAATGCTTATTCAGCAAAAATAAAGGCTTTTACCCAT 2802

QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221

Db 2803 TCTTTAGCTGTAGCTGGAAACCAAGGGGATTTGACGGGTGCCATTTATATCTCAACGA 2862

QY 222 AsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArg---Ala 240

Db 2863 AAT-----TCAATTGAAACCCCAAGTCCATAAAGATGCA 2895

QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260

Db 2896 TTAAAGGCGTACAAAGT-----TATCATCGAATTAATGCCAAACCAAGAGATCAA--- 2946

QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTyrGlnLys 280

Db 2947 -----TCTGCATACCTTTGTGTCGAAGTGTGTCGAAGCCAGATGATTAACAGT 3000

QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292

Db 3001 TGTTTACCTTTTCGCCAAACGACCTGGGATTTATCTCTCCCAAGAGAACCGTAAGCGTT 3060

QY 293 -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy 309

Db 3061 TCAGATTATACGGGGCTAACCGTATCAACCTTAATCCATGAATGAATGAAGCCAGTCT 3120

QY 309 sGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProIleAs 329

Db 3121 TGGTTTTTAAGAGAGGGTAT-CATTTTCTGCAACCAATATATTTGGTGGTATTTTGA 3179

QY 329 PheuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349

Db 3180 ATTCACACAAACAAATTTGATATCCGT-----GATATGACATTTCCCGCTTAT-- 3228

QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369

Db 3229 -----TTAAGATCAACAGAAAAACGGGATGATAG 3257

QY 369 aValGluAsnPheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387

Db 3258 CAGTGGCTCTTTTATCCAAAGCAAGATTTATGTTGTCATATCAACGTTATGAG----- 3309

QY 387 sGlnProIleGlyArg-----Le 393

Db 3310 -----GATGCCCGAGGCGTTAACTATGCAAGTGGGCTTTATTTTCGATGAACACCATAG 3362

QY 393 uLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413

Db 3363 AAAACAGCGGTAGTATTGAATATATTTACGAAATAAGAACAAAGCGGCATCAATTGA 3422

QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428

Db 3423 CAAAGCAGTGTTAAGTGTCTAATCAACAAACATCATCTTTGACAGTTTATATGCAACATAC 3482

QY 429 -----GlnAlaAsnTrp-----Tyr 429

Db 3483 GCATTGCAGTCTTTATTCCTTAATCCAAGTAAGATTGGCGCCCAACACGTGATAAACCTTA 3542

QY 429 rSerPhePheGlyValGlu----- 435

Db 3543 TTCATATCTATCATCTGATAGAAATGTTTATAAGAAAAAATAATATGTTGCAATTGAA 3602

QY 436 -----GlnAlaAsnTrp----- 439

Db 3603 TTTAGAGAAAAAATTCAACAAATTTGGCTTACTCATCAAAATGTTCTTCAATCTTGGTTT 3662

QY 440 -AspAsnPheThr-----LeuGluGlyGlyValArgValG 451

Db 3663 TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTTAACTCGACGTGTTCGCGTAC 3722

QY 451 uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLy 471

Db 3723 GGCAAAAGAGTATTTC-----GAGAAAGCTAATGAAACAAAGAAATGTTTACAA 3773

QY 471 sGln-----ProLeuProAspLeuGly----- 478

Db 3774 AAAACAACCTTACTTTATACCCAAACCAACAGTAGTGTTTTGTAGTACAAGATCATTTGTA 3833

QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492

Db 3834 TTATAAAGGTAACTCTCTTAATTACAGAGACTGTAAAGTCGGTTTAATTAAGGGAATAA 3893

QY 492 nTyrTyrPheThrProGlnHisLysLeuSerLeu----- 503

Db 3894 TTATTATTTCGACGACGCAATAATATGTCATTAGGGAATAACGTTGATTAGGTTTAGG 3953

QY 504 -----ThrAlaSerHisGlnGlu----- 509

Db 3954 TATTCGGTATGACGTATCTCGCAAAAAGCTAATGAATCAACTATTAGTGTGTGTAATTT 4013

QY 509 ----- 509

Db 4014 TAAAAATTTCTCTTGGAACTACTGGTATTGTCTATAAAACCAACGGAATGGCTTGATCTTTC 4073

QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522

Db 4074 TTATCGCCTTTCTACTGGATTTAGAAATCTCTAGTTTGTCTGAAATGTAT-----GGTTG 4127

QY 522 shiSValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542

Db 4128 GCGGTATGTGGCAATAATAGCGAGGTTTATGTAGGTAAATTTTAAAGCTGAAACATCTCG 4187

QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTy 562

Db 4188 TAACCAAGAGTTTGGTCTCGCTCTAAAGGGGATTTTGGTAATATTGAGATCAGTCATTT 4247

QY 562 rArgAsnArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsn-----AspG 578

Db 4248 TAGTAATGCTTATCGAAATCTTATCGCTTCTGCTGAAGAACTTATAAAAAATGGAACCTG 4307

QY 578 yArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn---Gl 597

Db 4308 AAAGGCCAATTATGATATCATATGCAAAAATGCAAAAATTTAGTTGGGTAAATATAAC 4367

QY 597 nSerGlyAlaAspPheTyrGlyValaGluGlyGluIleTyrPhe----- 611

Db 4368 TCGGCAATTAGATTTTATGGTTTATGGAAACGATTTCCCTACGGTGTGTGATGCAACATT 4427

```

QY 611 -----
Db 4428 TGCTTATAACCGAGTAAAGTAAAGATCAAAAAATCAATGCTGGTTCCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTGATGCCATTCAGCCAGC--CGTATATCATTTGGTTTAGGCTATGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCAAGTAATACITGGGGAATTAATACAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAs 653
Db 4605 AAATGAATTCCTAGGAAAACGT--GCATTGGGTAACAAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAGAAACTTACTCGGCA-----TGCATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGTATCGGTTATTACATGCGCAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
Db 4723 -----AATATTATGCTTCGATTAGGGATATATTAATTATTCAA 4760
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGTATGTTACTTGGGAAGCGGTGCTCAAAACAGACACAAGGTGCGGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATATCTCGTACGAGCATCA-----GGACGAAACTA 4868
QY 750 eThrGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTGAAATGAATTC 4893

```

RESULT 15

```

US-09-815-242-4030
; Sequence 4030, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4030
; LENGTH: 2442
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4030

Alignment Scores:
Pred. No.: 6,25e-09 Length: 2442
Score: 178.50 Matches: 180
Percent Similarity: 35.73% Conservative: 103
Best Local Similarity: 22.73% Mismatches: 279
Query Match: 4.42% Indels: 233
DB: 9 Gaps: 43

US-09-936-377-2 (1-758) x US-09-815-242-4030 (1-2442)
QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 374 AAATCAACCGCTCGCGC-ACCAGCAGCGACTGGTCTCGCCACGCGCCAGTCCACCGTC 432
QY 61 IleSerGly-----AspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu 77
Db 433 ATCGACATGCCCAACTGGAGGAGCTGGCGCAGGGTCCGACAGCTGGCCACCGTCTG 492
QY 78 -AspGlyValProGlyIleHisAlaSer-----GlnTyrGlyGlyGly 91
Db 493 GCCAAGCGGTGCGCGGATGTCGATTCAGCGCCACCATCACCGAATAC----- 543
QY 92 AlaSerAlaProValIleArgGlyGlnThr-----GlyArgGlyLeuValLeuAsn 109
Db 544 -----GGCCAGACCTCGCGGGCGCAGCATGCTGGTGTATGGTC 582
QY 110 HisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp 129
Db 583 GACGGCGTGGCTGAACACCAACCGCGACTCTCGCGCAACCTGGCC---AACATCGAC 639
QY 130 ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSer 149
Db 640 CCGCGCTGATCGAGCGGATCGAGTCACTCGTGGCAGC---AGCGCCATCTACGGCAGC 696
QY 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu 169
Db 697 GCGCCACCGCGGGATCATCTCCATCACCC-----ACCGTCCCGCC 738
QY 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn---LeuGluLysLeu 188
Db 739 GCGCGC---GAGAACCGCGGGAAACCGCGCTCAGCGCCACCTCGCGCTGACCCGCTG 795
QY 189 ThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu 208
Db 796 GCGAGCGAT-----GGCCTCGCGCGCCAGTTC----- 822
QY 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAsp 228
Db 823 -----CAGCAATA-CTTCGCGCGCTCCCTGGG 848
QY 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 248
Db 849 GCGCGCTCGACTATTTCGTTCTGACTTCGCGCACCCGCCA-----CGTCCGCGCTTC 896
QY 249 TyrArgArgThrTyrSerAspArg-----AspGlnTyrGlyLeuProAlaHis-Se 266
Db 897 CTACGACGCCCATGCGGACCGCATCGCCCGGAAACCCAGCCAGCGGACCTGTTTCGACTC 956
QY 266 rHisGluTyr-----AspAspCysHisAlaAspI1 276
Db 957 GAACGCTCAACATCGCGCGCAAGTCTGCGCTCGCATCGACAG----- 1002
QY 276 eIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuTh 296
Db 1003 -----AACGAGCGGTGAGTCTCGCTCGCCCTCAGCCACTACGACGC 1040

```



```
QY 57 -----SerAspLysIleIleSer 62
Db 202 GATATGCGGATGCTGGATATCCCGCAGGTGTCTAATACGGTTAGCGATCAGGTTCTGGAA 261
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 262 -----AACCAGAAATGCACCAACGCTGTGATGAGCGCTTTTATAACGTCAGTAAC 309
QY 83 ---IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThr 101
Db 310 GTGTGACAGCAATACATTAGGCGGAGCTCAGATGCTTTTGACCGCGTGGGTTTGGC 369
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGlnThrGlyAspMetAlaAspPheSer 121
Db 370 GCAAAACCGG----- 378
QY 122 ProAspHisAlaIleMetValAsp-----ThrAla 131
Db 379 ---GATGGCTCCATCATGACCAACGCTGTGCGAACCGTACTTCTCGTAGTTTCAACGCC 435
QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db 436 GCACAGAGCGTGTGGAGTCTTAAGCGCCGCCCTCCAGCTGTATGCGCATTTCTCGAT 495
QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly 171
Db 496 CCTGCGGAGCTGATTACGTCGTGACC---AAGCGCCGGAAGAAACATTCCATGTTGTCG 552
QY 172 ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly 191
Db 553 GTTTTCAGCC-----ACCTCTCC 570
QY 192 GlyIleAsnIleGlyLeuGlyLys-----AsnPheValLeuHisThrGluGly 207
Db 571 AGTTTGTGCGGCACTGGCACTTGATATACAGGTCCCATGAGGCACTCAGCTG 630
QY 208 LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db 631 GCGTATCGCCVTACCGGGGAAGTG----- 654
QY 228 AspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db 655 -----CAGATGAA 663
QY 248 PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db 664 GATTACTGGCGAAACTTCGGTAAAGAGCGCAGTACATTATTGCCCCCTCA----- 714
QY 268 GluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr 287
Db 715 -----CTCACCTGGTTGGTGAT-----AATGCAACAGTA 744
QY 288 LeuGlnLeuTyrProHis----- 293
Db 745 ACCATGCTCTATTCCCATCGGACATAAABCTCATTCGATCGTGGAACGATTTTCGAC 804
QY 294 LeuLeuThrGluGluAspValAsp-----TyrAspAsnProGlyLeu 307
Db 805 CTTACGACGAAACACACCGCCGTAACCGTTGATCGAAAAATACGTTTTCGAAACCG----- 858
QY 308 SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyr 327
Db 859 -----TTTAATTATACAGAT-----GGTCAGTCCGAT 885
QY 328 IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGly 347
Db 886 CTGGCGCACTCAACGAGAGATATCATCTCAATAGCCAGTGCAGCCGCGCTTT----- 939
QY 348 PheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGly 367
Db 940 -----GATTACAGCTACGCCAG----- 957
```

```
QY 368 AspAlaValGluAsnPhePheAsnAsnGln----- 377
Db 958 -----GATAAATACAGCGATTAATCAGCGCGTGTACCGCGTATGATGCACACGACA 1008
QY 378 ---ThrGlnAsnAlaArgIleGluLeu-----ArgHisGlnPro 389
Db 1009 GGAACACTGACACCGCGTGTGTGATGCAACTCAGCGATCTACCCACGCGTATGCATGCTACT 1068
QY 390 IleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db 1069 CGTCGGGATCTGCAAGGAAT-----GTTGATATTGCCGATTTCTAATATGAGATCTGT 1122
QY 410 SerAlaThrSer-----GluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db 1123 GGTGGGTGTATATGAATATTATGATCTTCTCGGTACAGATATGATTCGCTGTAAGAAA 1182
QY 426 ValGlnHisTyrSerPhePheGly----- 433
Db 1183 GCTAAAGATTTCATATATATACACCCCTGTTTATGTTAATACCGCAAAATGTACACCGTT 1242
QY 434 -----ValGluGlnAlaAsnTyr----- 439
Db 1243 TCGCGCTCGACACGGATCAGACGATCAACAGAGAGAACTACTCAGCTTATGCACAGGAC 1302
QY 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
Db 1303 CGGCTCTATCTGACCGGATAACTGATTGCCGTCGCCGGATCCGCTATCAGTATTACACG 1362
QY 455 AlaSerIleArgTyrAspLysAlaLeu---IleAspArgGluAsnTyrTyrLysGlnPro 473
Db 1363 CAATATCGGGTAAAGCGCTCCTTTTAATGTCATTAATCTGACACCGCGGATGAACAATGG 1422
QY 474 LeuProAspLeuGlyValAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyr 493
Db 1423 ACGCCCAAACTGGGGTAGCTACAAACTGACGCCATCGGTATCTTATTGGCCAATTAT 1482
QY 494 -----TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 1483 TCGCAAAACATTTATGCGCGAG-----TCGTCAATTGCGCAGTACATATGGCGATCTT 1533
QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 1534 CCACACGAA-----TCATCTAATGCTTACGAAGTC 1563
QY 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 1564 GGGCCAAA-----TTCGAGCTATTTCGATGGTATCACC 1596
QY 552 GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr 571
Db 1597 GCAGAT-----ATTGCGCTGTTTGATATCCATAAACGTAATGTGTGTAT 1641
QY 572 AlaGlnThrLeuAsnAsp-----GlyArgq---GlyProLysSer 583
Db 1642 ACCGAAAGTATTGTGATGAAACCATGCCAAAACGGCAGCGCGGTTCGTTCAAGAGG 1701
QY 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 1702 GTAGAAGTCGACCTTCGGGAGCATTAACGAAACATTATATCATATGCGCAGCTACCGC 1761
QY 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 1762 TATACCGATGCAAGGTTCTGGAAGATCCT----- 1791
QY 624 AspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 1792 GATTATGACGGGAACCATTTGCCGAATGTTCTCTCGTCATACCGGT----- 1836
QY 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 1837 -----TCGTATTCTCTGACCTATGATATTTCATTAACATGCCAGCAATAACACTGACG 1890
QY 664 Leu-----GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
```



```

Db      1891  TTTGGCGGTGGGACATGTTGTAAGCGTCGTTCCGCAACCAAT-----GGG 1938
Qy      680  LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db      1939  GCTGACTATTAT-----CTGCCT 1956
Qy      700  GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrp 719
Db      1957  GGCTATTTCGTTGCCGATGCTTCGCCGCATACAAAATGAAAATGCAGTAT---CCGGTC 2013
Qy      720  AsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSer 737
Db      2014  ACTCTGCAATTAACGTCAAAACCTGTTTGATATAAAGGTATTACACTCTTCC 2067

RESULT 17
US-10-085-959-152
; Sequence 152, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1e1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 152
; LENGTH: 7306
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-085-959-152

Alignment Scores:
Pred. No.:      4.15e-08      Length:      7306
Score:          177.50      Matches:      167
Percent Similarity: 31.47%      Conservative: 103
Best Local Similarity: 19.46%      Mismatches:  283
Query Match:      4.40%      Indels:      305
DB:              12          Gaps:        41

US-09-936-377-2 (1-758) x US-10-085-959-152 (1-7306)
Qy      6  LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db      5128  CTCAAAGGTCGGCGCTCTTTTCACTGCTTTTTCGGCAGCGATGATT-----CAT 5178
Qy      26  GluThrGluGlnSerValGlyLeuGlu-----ThrValThrValValGlyLys 41
Db      5179  GCACAGACTCTGTAAACGACCAAGATGGCGAAACAATCACTGTACAGCAGATGCAAAAT 5238
Qy      42  SerArgProAlaThrSerGlyLeu-----LeuHisThrSerThrAla----- 56
Db      5239  ACCGCACTGAGGCAACCGATGGTTATCAACCTCTGAGCACCTCCAGCGGACATTAAACC 5298
Qy      57  -----SerAspLysIleIleSer 62
Db      5299  GATATGCCGATGTGGATATCCCGAGGTGTCATACCGGTTAGCGATCAGGTTCTCGAA 5358
Qy      63  GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db      5359  -----AACCAAGATGGCAACACGCTGGATGAGCGCTTTATAACGTCAGTAAC 5406
Qy      83  ---IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThr 101
Db      5407  GTGTTACAGCAATACATATAGGGGGGACTCAGATGCTTTTGTACCCGCTGGGTTGGC 5466
Qy      102  GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
```

```

Db      5467  GCRAAACCGG----- 5475
Qy      122  ProAspHisAlaIleMetValAsp-----ThrAla 131
Db      5476  --GATGGCTCCATCATCATGACCAACCGGTCGCGAACCGTACCTCTCGTAGTTTCAACGCC 5532
Qy      132  LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db      5533  GCAACAGAGCGTGTGGAGTGTCTAAAGGCGCGCTCCACGCTGTATGGCATTTCTCGAT 5592
Qy      152  ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly 171
Db      5593  CCTGGCGGACTGATTAACTGCTGACCC--AAGCGCCGGAAACAACTTCATCCATGGTTCG 5649
Qy      172  ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly 191
Db      5650  GTTTCAGCC-----ACCTCTCTCC 5667
Qy      192  GlyIleAsnIleGlyLeuGlyLys-----AsnPheValLeuHisThrGluGly 207
Db      5668  AGTTTTCGTGGCGGCACTGGGCAACTTGATATCATCAGGTCCCATTTGAAGGCACTCAGCTG 5727
Qy      208  LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db      5728  GCGTATCGCCTTACCGGGGAGTC----- 5751
Qy      228  AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db      5752  -----CAGGATCAA 5760
Qy      248  PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db      5761  GATTACTGGCGAAACTTCGGTAAAGCGCAGTACATTTATTGCCCGGTCA----- 5811
Qy      268  GluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyr 287
Db      5812  -----CTCACCTGGTTTGGTGAT-----AATGCAACAGTA 5841
Qy      288  LeuGlnLeuTyrProHis----- 293
Db      5842  ACCATGCTCTATTCCTCCATCGGACTATAAACTCCATTTCGATCGTGGACGATTTTCGAC 5901
Qy      294  LeuLeuThrGluGluAspValAsp-----TyrAspAsnProGlyLeu 307
Db      5902  CTTACGACGAACACAGCCCGTAAACGTTGATCGAAAAATACGTTTTCAGCAACCG----- 5955
Qy      308  SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrp 327
Db      5956  -----TTTAATATTACAGAT-----GGTCAGTCCGAT 5982
Qy      328  IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGly 347
Db      5983  CTGGCGCAACTCAACGCAAGATATCATCTCAATAGCCAGTGGACAGCGCGCTTT----- 6036
Qy      348  PheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGly 367
Db      6037  -----GATTACAGCTACAGCCAG----- 6054
Qy      368  AspAlaValGluAsnPhePheAsnAsnGln----- 377
Db      6055  -----GATAAATACAGCGATAATCAGCGCGGTGTACCGCGTATGATGCAACGACA 6105
Qy      378  ---ThrGlnAsnAlaArgIleGluLeu-----ArgHisGlnPro 389
Db      6106  GGAACACTGACACGCGCTGTTGATGCAACTCAGGGATCTACCCAGCGCTATGATGCTACT 6165
Qy      390  IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db      6166  CGTGGCGGATCTGAAGGGAAT-----GTTGATATTCCCGGATTTCTATAATGAGATTCTG 6219
Qy      410  SerAlaThrSer-----GluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db      6220  GGTGGGGGTGCATATGATATTATGATCTTCTCGGTACAGATATGATTCGCTGTAAAAA 6279
```

```
QY 426 ValGlnHisTyrSerPheGly----- 433
Db 6280 GCTAAGATTTCATATATACACCCCGTTTATGGTAATACCAGCAATGTACACGGTT 6339
QY 434 -----ValGlnAlaAsnTrp----- 439
Db 6340 TCGCGCTCGGACGCGATCAGACGATCAAAAGAGAACTACTCAGCTTATGCACAGGAC 6399
QY 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
Db 6400 GCGCTCTATCGACCGAATACTGATTCGCGTCCCGGATCCGCTATCAGTATACAG 6459
QY 455 AlaSerIleArgTyrAspLysAlaLeu---IleAspArgGluAsnTyrTyrLysGlnPro 473
Db 6460 CAATATCGGTAAAGCGCGCTTTTAATGTCAATACTGACAGCCGCGATGAACAATGG 6519
QY 474 LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrp 493
Db 6520 ACGCCCAACTGGGGTTAGTCTACAAACTGACGCCATCGGTATCTCTATTGGCCAATAT 6579
QY 494 -----TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 6580 TCGCAACATTTATGCCGCG------TCGTCAATTCGACGCTACATTTGGCGATCTT 6630
QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 6631 CCACAGAA-----TCATCTAATGCTTACGAAGTC 6660
QY 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 6661 GGGCAAAA-----TTCGAGCTATTCGATGGTATCAC 6693
QY 552 GlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr 571
Db 6694 GCAGAT-----ATTGCGCTGTTTGATATCCATAAAGCTAATGTGTGTAT 6738
QY 572 AlaGlnThrLeuAsnAsp-----GlyArg---GlyProLysSer 583
Db 6739 ACCGAAAGTATTTGGTGAAGAACCATCGCCAAACCGGCGCGCTTCGTTCAAGAGGG 6798
QY 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 6799 GTAGAAGTCGACCTTGGGAGCAATTAAGTAAACATTAAATATCATTCGCGCTACGCG 6858
QY 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 6859 TATACCGATGCAAGGTTCTGGAAGATCCT----- 6888
QY 624 AspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 6889 GATTATGAGGAAACCATTCGGAATGTTCTCGTCAATACCGT----- 6933
QY 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 6934 -----TCGTATTCTCGACCTATGATTTATCAACATGCCAGCAATACACTGACG 6987
QY 664 Leu-----GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db 6988 TTTGGCGTGGCGACATGGTAAAGCGCTGTCGGCAACCAAT-----GGG 7035
QY 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db 7036 GCTGACTATTAT-----CTGCCT 7053
QY 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrp 719
Db 7054 GGCATTATTCGTTGCCGATCGCTTCGCCCATACAAATGAATTCAGATAT---CCGCTC 7110
QY 720 AsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSer 737
Db 7111 ACTCTGCAATTAAACGCTCAAAACCTGTTTGTATAAAACGTATTATACACCTCTTCC 7164
```

RESULT 18

```
US-10-238-075-261
; Sequence 261, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.B.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol.
; FILE REFERENCE: E.coli, and biological uses of these polynucleotides and of th.
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 261
; LENGTH: 7315
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-261

Alignment Scores:
Pred. No.: 4,16e-08 Length: 7315
Score: 177,50 Matches: 167
Percent Similarity: 31,47% Conservative: 103
Best Local Similarity: 19,46% Mismatches: 283
Query Match: 4,40% Indels: 305
DB: Gaps: 41

US-09-936-377-2 (1-758) x US-10-238-075-261 (1-7315)
QY 6 LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db 5137 CTCAAAGTGGCGGCTCTTTTCACTGCTTTTTCGCGACCGATGATT-----CAT 5187
QY 26 GluThrGluGlnSerValGlyLeuGlu-----ThrValThrValValGlyLys 41
Db 5188 GCAACAGACTCTGTAAACGACCAAGATGGCGAAACAATCACTGTACAGCAGATGCAAT 5247
QY 42 SerArgProArgAlaThrSerGlyLeu-----LeuHisThrSerThrAla----- 56
Db 5248 ACCGCAACTGAGGCAACCGATGGTTATCAACCTCTGAGCACCTCCACGGCGACATTAA 5307
QY 57 -----SerAspLysIleIleSer 62
Db 5308 GATATGCCGATGCTGGATATCCCGAGTGTCATAACGGTTGATCGATCAAGTTCTGGA 5367
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProgly 82
Db 5368 -----AACCGAATGCCGACCAACGCTGGATGAGCGCTTTATACGTCAGTAAC 5415
QY 83 ---IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThr 101
Db 5416 GTGTACAGACCAATACATAGGCGGACTCAGGATGCTTTTGTACGCGCGTGGTTTGGC 5475
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 5476 GCAAAACCGG----- 5484
QY 122 ProAspHisAlaIleMetValAsp-----ThrAla 131
Db 5485 ---GATGGCTCCATCATGACCAACCGCTGCGAACCCTCTCTCTAGTTTCAACGCC 5541
QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db 5542 GCAACAGAGCGTGTGAAGTGTCTAAAGGCCCGCCCTCCACGCTGATGGCATCTTCGAT 5601
QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly 171
Db 5602 CCTGGCGACTGATTAACGTCGTGACCC---AAGCGCGCGGAAAAACAATTCATCGTTCG 5658
QY 172 ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly 191
```

Db 5659 GTTTCAGCC-----ACCTCCTCC 5676
Qy 192 GlyIleAsnIleGlyLeuGlyLys-----AsnPheValLeuHisThrGluGly 207
Db 5677 AGTTTGTGGCGGCACTGGCACTTGATATACAGGTCCCATGAGGCACTCAGCTG 5736
Qy 208 LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db 5737 GCGTATCGCCATTACCGGGAAGTG-----5760
Qy 228 AspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db 5761 -----CAGATGAA 5769
Qy 248 PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db 5770 GATTACTGGCGAAACTTCGGTAAAGAGCGCAGTACATTAATTGCCCCGTCA-----5820
Qy 268 GluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr 287
Db 5821 -----CTCACCTGGTTTGGTGAT-----AATGCAACAGTA 5850
Qy 288 LeuGlnLeuTyrProHis-----293
Db 5951 ACCATGCTCTATCCCATCGGAGCTATAAACTCCATTTCGATCGTGGAAACGATTTTCGAC 5910
Qy 294 LeuLeuThrGluGluAspValasp-----TyrAspAsnProGlyLeu 307
Db 5911 CTTACGACGAACACCGCCGTAACCGTTGATCGAAAAATACGTTTTGACGAACCG-----5964
Qy 308 SerCysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyr 327
Db 5965 -----TTTAATATTACAGAT-----GGTCAGTCCGAT 5991
Qy 328 IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGly 347
Db 5992 CTGGCGCAACTCAACGCAAGATATCATCTCAATAGCCAGTGCAGCGCGCTTT-----6045
Qy 348 PheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGly 367
Db 6046 -----GATTACAGCTACAGCCAG-----6063
Qy 368 AspAlaValGluAsnPheAsnAsnGln-----377
Db 6064 -----GATAAATACAGCATATACAGCGCGTGTACCGGTATGATGCAACAGACA 6114
Qy 378 ---ThrGlnAsnAlaArgIleGluLeu-----ArgHisGlnPro 389
Db 6115 GGAACACTGCACGCGGTGTGATGCAACTCAGGGATCTACCCAGCGTATGATGCTACT 6174
Qy 390 IleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db 6175 CGTGGGATGCAAGGAAT-----GTTGATATGCGGATTCCTATAATGAGATTCTG 6228
Qy 410 SerAlaThrSer-----GluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db 6229 GGTGGGTGTATATGATATATGATCTTCTGCGTACAGATATGATTCGTGTAAATAA 6288
Qy 426 ValGlnHisTyrSerPheGly-----433
Db 6289 GCTAAAGATTCAATATATACAACTGTTTATGGTAAATACAGCAAAATGTACAACGGTT 6348
Qy 434 -----ValGlnAlaAsnTyr-----439
Db 6349 TCGGCGTCCGACGAGTACAGCATCAACAGAGAGAACTACTCAGCTTATGCACAGGAC 6408
Qy 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
Db 6409 GCGCTCTATCTGACCGATAATCGATTCCGTCGCGGAGTCCGTTATCAGTATTACAG 6468
Qy 455 AlaSerIleArgTyrAspAlaLeu---IleAspArgGluAsnTyrTyrLysGlnPro 473
Db 6469 CAATATCGGGTAAAGCGCGCTCTTTTAATGTCAATACTGACAGCGCGGATGAACATGG 6528

Qy 474 LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyr 493
Db 6529 ACGCCAAACTGGGTTAGTCTACAAACTGACCCATCGGTATCTTATTTGCCAATTAT 6588
Qy 494 -----TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 6589 TCGCAAAACATTATGCGCGCAG-----TCGCAATTGCGCAGTACATATGGCGATCTT 6639
Qy 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 6640 CCACCAGNA-----TCATCTAATGCTTACGAGTTC 6669
Qy 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 6670 GGGGCAAAA-----TTCGAGCTATTTCGATGATCACC 6702
Qy 552 GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr 571
Db 6703 GCAGAT-----ATTGCGCTGTTTGATATCCATAAACGTAATGTGTGTAT 6747
Qy 572 AlaGlnThrLeuAsnAsp-----GlyArg---GlyProLysSer 583
Db 6748 ACCGAAAGTATGTTGATGAACCATCCCAAAACGGCAGCGCGTTCGTTCAGAGGG 6807
Qy 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 6808 GTAGAAGTCGACCTTGGCGGAGCATTAACCTGAAACATTAAATATCATTTGCCAGCTACGCG 6867
Qy 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 6868 TATACCGATCAAAAGTTCTGGAAGATCCT-----6897
Qy 624 AspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 6898 GATTATGCGAGGAACCATTCGCGAATGTTCTCGTCATACCGT-----6942
Qy 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 6943 -----TCGTATTCTCGACCTATGATATTCAATCATGCCAGGCAATAACACACTGACG 6996
Qy 664 Leu-----GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db 6997 TTTGCGGTGCGGACATGTTGTAAGCGCTGTTGCGCAACCAAT-----GGG 7044
Qy 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db 7045 GCTGACTATTAT-----CTGCT 7062
Qy 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyr 719
Db 7063 GGCTATTTGTTGCGGATCGCTTCCGCGCATACAAAATGAAATTCAGTAT---CCGGTC 7119
Qy 720 AsnTyrTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSer 737
Db 7120 ACTCTGCAATTAAACGTCAAAACCTGTTTGATAAAACGATTACACCTCTTCC 7173

RESULT 19

US-10-085-959-24/c
; Sequence 24, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 24

; LENGTH: 6732

; TYPE: DNA

; ORGANISM: Escherichia coli

US-10-085-959-24

Alignment Scores:

Pred. No.: 4, 73e-08 Length: 6732
 Score: 176.50 Matches: 148
 Percent Similarity: 32.67% Conservative: 97
 Best Local Similarity: 19.73% Mismatches: 294
 Query Match: 4.37% Indels: 211
 Db: 12 Gaps: 32

US-09-936-377-2 (1-758) x US-10-085-959-24 (1-6732)

QY 27 ThrGluGlnSerValGlyLeu-----GluThrValThrValValGlyLysSerArg 43
 Db ACCAGAGTCCGTAGCGCTCAGGAGAGTGACAGCGTACCGTATGGTCCAGTCCGTA 1840
 QY 44 ProArgAlaThrSerGlyLeuLeuHis-----ThrSerThrAlaSerAspLysIleile 61
 Db TCATCGACGACACCAACCGCTTCTCGATCAACCCACCATGAAGCCCTCGATAAA----- 1786
 QY 62 SerGlyAspThrLeuArgGlnLysAlaValAlaLeuGlyAspAlaLeuAspGlyValPro 81
 Db -----CAGAAATGCTCGCTCAGGCATTAAAGTCTGCTCCCC 1753
 QY 82 GlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGln 100
 Db GCGGTGGTCTGCAAAAGTCAGGTAGCGCAACGAAGAAGTCAAGTTCGTGCTTT 1693
 QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db GATAGTCGTAGGTCCGCTCTATTTCGACGGTGTGCCCATTTATCTCCCTATCACGCG 1633
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140
 Db AACCTCGATCTGGCGGATCTGACCAACAAATTTGGGG---GCAGTTGAAGTTTCCAAA 1576
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db GGGTATTGCTGCTCTTCAGGGGCTAATCAGATGGGAGGCCATTAAATATCACACC 1516
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgIleu 180
 Db CAGAAAG---CCGACTAAACCGCTGGAAGCCAGTCTGGGATATCGTCAGGCGATGGA 1459
 QY 181 SerSerGlyAsnLeuGluLysLeu---ThrSerGlyGlyIleAsnIleGlyLeuGlyLys 199
 Db ACCCAGAACATGCCTACGATATGCTTCTATTGGCCGCGCAGCAGCGATCTGGGG--- 1402
 QY 200 AsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArg 219
 Db -----TATTTCAGAGTCAGCGGTAGCCAGCTAAAGCAGGATTTCTCGGCTCGCG--- 1351
 QY 220 TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
 Db -----CATGGTGTAAATAAT 1336
 QY 240 AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
 Db -----CATATTGAGGCAACACCGCAAGATGATTATTCATCGCTGATGATAACCGCGCAT 1276
 QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
 Db -----GTGAGGCTGGTTTACACCCACCGAAACGATGAA----- 1240
 QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysAsp 299
 Db -----TACATTCAGCTTACATTAAGCAGGAT 1213
 QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319

Db 1212 GGTGAAAAGATAACCCGCCATACAGT----- 1186
 QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
 Db -----CAGAAATCAGGTAC----- 1159
 QY 340 GluTrpLysGlnPro-----PheProGlyPheGluAlaLeu 351
 Db -----TGGCAGTGGCCAGAGTATGACAAAGAAAGTTTATTATTCAGGGAACGACCACTA 1102
 QY 352 -----ArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAla 369
 Db AACGATCGTTTACCTGAAAAAGTCGCTGATCAC-----GACACC 1060
 QY 370 ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro 389
 Db TTTGAAATACGCTA-----ATGATGTACAACTCGCTGGGTGATTTGAAAAATAAA--- 1009
 QY 390 IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
 Db -----AAAGGCAGCTACAGC---CATTTATCCGATTACAGCGACGTCGCGG 964
 QY 410 SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGln----- 427
 Db TTACAACCTGGCAGCGCATGTCGTGAAAAACGATCTGCTGCTTTCGCGTTAACTGGAAA 904
 QY 428 -----HisTyrSerPhePheGlyValGluGlnAla 437
 Db GATGACGTACACCGGAAAAAGGTGCGCGCGCACCGGCTTACGATCGCTATGAAGATCGT 844
 QY 438 AsnTrp-----AspAsnPheThrLeuGluGly 446
 Db ACCTGCTGCTCGCCAGTGAATATCATGGGCTGCTGCGGATAATGTCGATCGTGGCT 784
 QY 447 GlyVal-----ArgValLysGlnLysAla 455
 Db GGAATCAGCTATGACTGCGCGGATAGCTAGAGCGGAAAAACATGAGAAAGATGCGACT 724
 QY 456 SerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuPro 475
 Db ATCAACCATTTATGACGAC----- 706
 QY 476 AspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPhe 495
 Db -----AACATCAGTCAGCTTTTAACTGGCAGGTGATGGGAAAAATACCACTTT 658
 QY 496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGln 515
 Db GCCAATGAAGCACCGCTGGCGCTTTCGTACTATGACCGCACACGCTTTCCGACGCTGAAA 598
 QY 516 GluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHis 535
 Db GAAACGTATACCGCTCCAAA---CCTGGTATACACAGATAGCAGTATGTTAACCGCGAG 541
 QY 536 LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu---GlyTyrGluGlyAspArg 554
 Db CTCAAGCCGGAACCGCGCGGGTGGATTAACTGGAAATGGTGCCTTCCAGCAGCAGAC 481
 QY 555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThr 574
 Db TGGGATTGAGGTACCGCTTTTACTATAACCGGTGAGTATGATGCGCATCTCTCGCAC--- 424
 QY 575 LeuAsnAspGlyArgGlyProLysSerIleGluAspSerGluMetLysLeuValArg 594
 Db -----AATATCATGCTCCGATACC-----ATTCAAAAT 397
 QY 595 TyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThr 614
 Db CAGAACACGCGCACCGGTGATTACAGCGGTCTGGATGCGGTATTAAAGGGAAAAATCAGC 337
 QY 615 ProArgTyrArgIleGlyValSer-----GlyAspTyrValArg----- 627

QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHis----- 293
Db 5821 GGTGAT-----AATGCAACAGTAACCATCTATTCCCATCGGACTATAAACTCCG 5874
QY 294 -----LeuLeuThrGluGluAspValAsp----- 301
Db 5875 TTCGATCGTGAACAGATTTTCGACCTTACACGAAACACGCCGTAAACAGTTGATCGAAAA 5934
QY 302 -----TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
Db 5935 ATACGTTTTCGCAACG-----TTTAATATTACAGAT----- 5967
QY 320 HisAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db 5968 -----GTCAGTCGATCTCGCCCACTCAACGCAAGATATCATCTCAATAGC 6015
QY 340 GluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db 6016 CAGTGGACAGCGCGCTTT-----GAT 6036
QY 360 TyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGln----- 377
Db 6037 TACAGCTACAGCCAG-----GATAAATACAGCGACAATCAGGCTCGC 6078
QY 378 -----ThrGlnAsnAlaArgIleGluLeu----- 385
Db 6079 GTTACCGCGTATGATCAACAGCAGCAACGCTGACCGCGTGTGTGATGCAACTCAGGGA 6138
QY 386 -----ArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyr 401
Db 6139 TCTACCCAGCGTATGCAATCTACTCGTGGGATCTGCAAGGAAT-----GTTGATATT 6192
QY 402 LeuGlyGlnLysSerAlaLeuSerAlaThrSer-----GluAlaValLys 417
Db 6193 GCTGGTCTTATAATGAGATTCCTGGTGGGTGTCATATGAAATATTATGATCTCTCGC 6252
QY 418 GlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAla 437
Db 6253 ACAGATATGATTCCTGTAAACGCTAAAGATTCAATATCTACAAACCCCGTTTATGGC 6312
QY 438 AsnTyrAspAsnPheThrLeu-----GluGlyGlyValArgValGluLysGln 453
Db 6313 AATACCGCAAAATCAACACGTTTCGGCGTCGACAGTATGATCAGACATCAACAGGAG 6372
QY 454 LysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsn----- 468
Db 6373 AGCTACTCAGCTTATGACAGAGATCGCTCTATCTGACCGATACTGGATTGCCGCTGCC 6432
QY 469 -----TyrTyr----- 470
Db 6433 GGGATCGCTATCAGTATTACACGATGATCGGGTAAAGCCGCTCTTTAATGTCAAT 6492
QY 471 -----LysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg 485
Db 6493 ACTGACAGCCGCGATGAACATGACGCCCAACTGGGGTTAGTCTACAACTCAGCGCA 6552
QY 486 SerPheAlaLeuSerGlyAsnTyr-----TyrPheThrProGlnHisLysLeuSerLeu 503
Db 6553 TCGTATCTCTATTGTGTAATATTTCGCAACATTTATGCGCAA-----TCGTCA 6603
QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db 6604 ATTGCCAGTACATCGGAGATCTCCACCGAA----- 6636
QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db 6637 ---TCATCTAATGCTTACGAAGTCGGGCAAAA----- 6666
QY 544 IleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg 563
Db 6667 TTCGAGCTGTTCGATGGTATTAACCGCAGAT-----ATTGCGCTCTTTGAT 6711
QY 564 AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp----- 577

Db 6712 ATCCATAAACGTAACGTGTGTATACCGAAAGTATTGGTGATGAAACCATCGCAAAACG 6771
QY 578 ---GlyArg---GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595
Db 6772 GCAGGCGCGTTCGTTCAAGAGGGTAGAAGTCACTTCGCGGAGCATTAACATAAC 6831
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
Db 6832 ATTAATATCATTCGACGTACGCTATACCGCATTAAGGTCTGGAAGATCCT----- 6885
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 6886 -----GATTATGCAAGGAAACCATTCGCGAATGTTCTCCGT 6921
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 6922 CATACCGGT-----TCGCTATTCTGACCTATGACATTCATAAC 6960
QY 656 AlaProArgValProAlaAlaArgLeu-----GlyPheHisLeu-----LysAlaSer 671
Db 6961 ATCCAGGCAATAACACACTGACGTTTGGCGGTGGTGCACATTGCGTAAGCGCTCG 7020
QY 672 LeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeu 691
Db 7021 GCACCAAT-----GGGCGTGACTATTAT----- 7044
QY 692 AlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 7045 -----CTGCCAGGCTATTTCGTTCCGATGCTTCGCGCATACAA 7086
QY 712 ArgAsnThrArgTyrGlyGluTyrAsnTyrTyrValLysAlaAspAsnLeuAsnGln 731
Db 7087 ATGAAATTCAGTAT---CCGTCACACTGCAATTAACCTCAAAAACCTGTTGATAAA 7143
QY 732 SerValTyrAlaHisSer 737
Db 7144 ACGTATTACACCTCTTCC 7161
RESULT 21
US-10-181-319-13
; Sequence 13, Application US/10181319
; Publication No. US20030135032A1
; GENERAL INFORMATION:
; APPLICANT: Lewis, Thomas A.
; APPLICANT: Paszczynski, Andrzej
; APPLICANT: Crawford, Ronald L.
; APPLICANT: Cortese, Marc S.
; APPLICANT: Sebati, Jonathan L.
; TITLE OF INVENTION: Compositions and Methods for Bioremediation
; FILE REFERENCE: IDAHI19440
; CURRENT APPLICATION NUMBER: US/10/181,319
; PRIOR FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: PCT/US01/02386
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/177,251
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 25801
; TYPE: DNA
; ORGANISM: Pseudomonas stutzeri
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 8300,8319,8338,8708,11735,11768,18459,18686,18703,18711,18720,19226,2007
; OTHER INFORMATION: N = unknown nucleic acid residue
US-10-181-319-13
Alignment Scores:
Pred. No.: 2,07e-06 Length: 25801
Score: 169.50 Matches: 168
Percent Similarity: 35.19% Conservative: 103

Best Local Similarity: 21.82% Mismatches: 251
Query Match: 4.20% Indels: 253
DB: 12 Gaps: 37

US-09-936-377-2 (1-758) x US-10-181-319-13 (1-25801)

QY 42 SerArgProAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIlelle 61
DB 12179 TCACAGCCGAGAAATCGAGCGCGG-CTGAAAGGTGGCCCGCAGCGTGGCGGTATC 12237
QY 62 SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly--Val 80
DB 12238 GATGGCTGGACCGCCGAGCAGTCCAGCATCAGCTCAACACACTGGAGAGCGATT 12297
QY 81 ProGlyIleHisAlaSerGlnTyrGly---GlyGlyAlaSerAlaProValIleArgGly 99
DB 12298 CCTGGTCTGTCATTCACAGCGGTTCGGGCAAGCAGGTATGAAATTCACCGCTCATCGCGGG 12357
QY 100 GlnThrGlyArgGlnLysValLeuAsnHisGlyGlnThrGlyAspMetAlaAsp 119
DB 12358 CTG-----ACGGCAACTTCAACAGC 12378
QY 120 PheSerProAspHisAlaIleMetValAsp----- 129
DB 12379 TTCTCCAGTTCAAGCTTGTGCTGCTCGATGGCTTCCACGCTGACAGCCCGAGGATT 12438
QY 130 -----ThrAlaLeuSerGln-GlnValGluIleLeuArgGlyProValThrLeuLe 146
DB 12439 CGAGAGTGGCATGTCATCGATCGATCGAGTCAATTCGCGCGCCGCAATCTACGCT 12498
QY 146 uTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro---G1 165
DB 12499 GTATGGCGTAATGCCAGCGCGGTGTG-----ATTGCCATCCACAGCTTCCGATGGA 12552
QY 165 uLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg----- 179
DB 12553 CGCAGCCCGAGAGCCAGCGGTGTCTGCCGAAGCGGCGAGCGGAACAGCGTGTATGCG 12612
QY 180 -----LeuSerSerGlyAsnLeuGlyLeuThrSerGlyGlyIleAsnIleGlyLe 197
DB 12613 GTTTGGCTTCAGCAGCGCTTTGGGAAGACGCTTGTACGGCGCGTA----- 12661
QY 197 uGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaVa 217
DB 12662 -----TCGGCAACTGTCTG-- 12676
QY 217 lProArgTyrArgAsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyG1 237
DB 12677 -----AGCCAGAGCGGCTTTCATCGACAACAC 12702
QY 237 nHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgThrTyrSerAspArgAr 257
DB 12703 CCACAG----- 12709
QY 257 gAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleI 277
DB 12710 -----GGCAACAAGCGCGACGATCGTGAR----- 12733
QY 277 eTrpGlnLysSerLeuIleAsnLysArgTyrGlnLeuTyrProHisLeuLeuThrG1 297
DB 12734 -----CAGAGAACCTG-----AACCTGGGCTCGCTGGGCGCGCGGCGCAACGGA 12783
QY 297 uGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAl 317
DB 12784 TGTGTCATGGCGTAT-----GC 12801
QY 317 aHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGlu-- 336
DB 12802 GCATCAGAGTACAAATGCGGCTCCCT-GTGGGCTCCCGCGCGCGCAAGGAAC 12860
QY 337 -----LeuArgAlaGluTrpLysGlnProPheProGlyPheGlu----- 349
DB 12861 AAGTCCCGTCCGGAACGCCCACTGGAACCGTTCTGAGGGCGCACACCTTGTCTTCAATG 12920

QY 350 -----AlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisaspGluLy 365
DB 12921 TYCAGCATGAATTTGCTCGGCTGCGTTCGATTCGGTAAACGGCCTGGACAGGATCAAG 12980
QY 365 sAlaGly-----AspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
DB 12981 GACAGGATTCAGCAGGACACTGACTTCATCCAGCCCATGTTCT-----TGCAGCTCGG 13034
QY 380 nAlaArgIleGluLeuArgHisGlnPro-----IleGlyAr 392
DB 13035 CGCAGCATCACCTGGCGCACACTCTCCAGAGTTCGTTGAGGAGCAGCTCGGGAG 13094
QY 392 gLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaTh 412
DB 13095 GCAGTTGGC---TGGCTGGTGC-TACGGCGGATCGCAGCAGCAACATCTGCACATAC 13150
QY 412 rSerGluAlaLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePh 432
DB 13151 CAGCAAGACCATGATGGG-----CTGTGGACATTCGCGCGGATCAGCAGCGGA-TA 13203
QY 432 eGlyValGluGlnAlaAsnTrpAsp-----AsnPheThrLeuGluGlyG1 447
DB 13204 CGCTGCACTGTTCACCCACTGGAACGTCCCTGTCGGCGGACTGTGTCATAGACGCGG 13263
QY 447 yValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArg1 467
DB 13264 AGCGCGGTTCAGCGCAACAGAGTGCAGTACGTCCGCAAGGGGCTACG----- 13312
QY 467 uAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPh 487
DB 13313 -----AGCATGAAAA----- 13324
QY 487 eAlaLeuSerGlyAsnTrpTyrPheThrPro-----GlnHisLysLeuSerLe 503
DB 13325 -----GGWGGACACACAGTTTCCAGCGCTCGCGTGCACACACAG-----AT 13368
QY 503 uThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHi 523
DB 13369 AACCGCAATCACCAGTATGTATGTAGTGCAGTGTGGCGTGGCGTGCAGCTGCGGCTTCAA 13428
QY 523 sValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAs 543
DB 13429 TGTGCTGGCGCGCGCTG-----GGTTATCTGCTTACGACCGGAGAA 13476
QY 543 nIleGluLeuAlaLeuGlyTyrGluGly-----AspArgTrpGlnTyrAsnLe 559
DB 13477 CTGTCGTATGAAACCGGTCTCAAGGCTGGCTTCTTGACAAGCGCATTCGCTATTCGCT 13536
QY 559 uAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyAr 579
DB 13537 GCGCGCTTAC-----CTCATGGAC----- 13555
QY 579 gGlyProLysSerIleGluAspAspSerGluMetLysLeu-----ValArgTy 595
DB 13556 -----ATCATGATCATGAGGTTCATGATGATGATGATGATGATGATGATGATGAT 13602
QY 595 rAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPr 615
DB 13603 CATCAGCAGCGCTCCACCGCAGCATCCAAAGGTCTCGAGTGTGAGTGTGAGTGTGAGTGT 13661
QY 615 oArgTyrArgIleGlyValSerGly-----AspTy 625
DB 13662 TCGTGGCGGCTCGACGCTCAAGGCGGCGGCTGGCTGGAAACACACAGCGCGCTTCGATCA 13721
QY 625 rValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAs 645
DB 13722 CTTTCG-----GATGGCGAGCGGACTATGACGCA 13754
QY 645 nArg---ProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLe 664
DB 13755 CCAAGACCGTTTCGCGC-----GGATCTCACCGGCCACCTCGGATCCGCTACGACGCG 13811

Qy 664 u-----GlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrT 683
 Db 13812 CCAGAGGCTGGTATGACCAAGCCAGCGTACCGCGCAGCAGCAAGGTATACVTCATCGCG 13871
 Qy 683 yrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisM 703
 Db 13872 CCACAGGCTATGAACGCAAC-----GGCTACGGCY 13901
 Qy 703 etLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGluThrAsnTyrV 723
 Db 13902 TGGTGAACCTGGTAGTGGTTACCAACG-----GGCAACTGGGAATCGCG 13949
 Qy 723 allysAlaAspAsnLeuAsnGln 731
 Db 13950 CCTACGCGCACACCGCAGCATCAG 13975

RESULT 22

US-10-114-170-76/c
 ; Sequence 76, Application US/10114170
 ; Publication No. US20030023075A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Blattner, Frederick R.
 ; Burland, Valerie
 ; Perna, Nicole T.
 ; Plunkett, Guy
 ; Welch, Rod
 ; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Pinckney Street
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: US
 ; ZIP: 53701-2113

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 8.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/114,170
 ; FILING DATE: 01-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: 09/453,702
 ; FILING DATE: 03-DEC-1999
 ; APPLICATION NUMBER: 60/110,955
 ; FILING DATE: 04-DEC-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J.
 ; REGISTRATION NUMBER: 27386
 ; REFERENCE/DOCKET NUMBER: 960296.95017
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (608) 251-5000
 ; TELEFAX: (608) 251-9166
 ; INFORMATION FOR SEQ ID NO: 76:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6737
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 76:

US-10-114-170-76

Alignment Scores:
 Pred. No.: 9,78e-07 Length: 6737
 Score: 164.50 Matches: 142
 Percent Similarity: 33.02% Conservative: 106
 Best Local Similarity: 18.91% Mismatches: 290
 Query Match: 4.08% Indels: 213
 DB: 14 Gaps: 31

US-09-936-377-2 (1-758) x US-10-114-170-76 (1-6737)
 Qy 27 ThrGlnGlnSerValGlyLeu-----GluThrValThrValValGlyLysSerArg 43
 Db 1899 ACCGAGGCGGTAGCGGCTCAGGAGCGACTCGTACCGGTATGGTCAGTCCGGTA 1840
 Qy 44 ProArgAlaThrSerGlyLeuLeuHis-----ThrSerThrAlaSerAspLysIle 61
 Db 1839 TCATCGACGACGACCGCTTCTCGATCAACCCACCATGAAGGCCCTGGATAAA----- 1786
 Qy 62 SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValPro 81
 Db 1785 -----CAGAATGTGCTCAGGCATTAAAGTGTGCTGCC 1753
 Qy 82 GlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal-----IleArgGlyGln 100
 Db 1752 GCGTGGTGTGCAAAAGTCAGGCGCGCAACGAGACAGTAAAGTTTCGTCGCTT 1693
 Qy 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 1692 GATAGTCGCGAGGTGCGGCTCTATTTCGACGTGTGCCATTATGTTCCTATGACGCG 1633
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLeuArg 140
 Db 1632 AACCTCGATCGCGGATTCCTGACCAACAATCTGGGG-----GCAGTCGAAGTTTCCAAA 1576
 Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 1575 GGGTATTCTGCTGCTTCAGGGACCTAATCAGATGGCGGAGCCATTAAATATCACACC 1516
 Qy 161 GlyLysIleProGluLysMetProGluAsnGlyVal----- 172
 Db 1515 CAGAAG---CCAACAACCTCTGGAAGCAAGTCTGGGATATCGCAGGAGTGGAGCGGT 1459
 Qy 173 -----SerGlyLysLeuGly--- 177
 Db 1458 AGCAGGACAATGCTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399
 Qy 178 LeuArgLeuSerSerGlyAsnLeuGlyLysLeuThrSerGly-----GlyIleAsn 194
 Db 1398 TTGCAAGTCAGCGGTAGCGCAGCTAAGCAGGAGTTCCTCGGCTCCGCGTGGTAAAT 1339
 Qy 195 IleGlyLeuGly-----LysAsnPhe 201
 Db 1338 AATGATATTGACGCGCAACACGCAAGATGATTAATTCATCGGTGATGATGATGATGATGAT 1279
 Qy 202 ValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg 221
 Db 1278 ATTGTGAAGCTCGGATTTACACCGTGAACCGATGAATACACTGACTTACATTAAAG 1219
 Qy 222 AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaVal 241
 Db 1218 CAGGATGGTGAAGAAAGATAACCGCCATACAGCGGAATATGGTCAAAAATCACGCTAC 1159
 Qy 242 LeuGlyTyr-----ArgLysArgPheTyrArgArgThrTyrSerAspArg 257
 Db 1158 TGGCAGTGGCCAGAGTATGACAAAAGAAAGTTTTTATTATCAGGGAACCAACCACTAAC 1099
 Qy 258 AspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIle 277
 Db 1098 GATCGTTTACCTGAAAGTCGGCTGTATCGC---GACACCTTTGAAAATACGCTGATG 1042
 Qy 278 TrpGlnLysSerLeuIleAsn---LysArgTyrLeuGlnLeuTyrProHisLeuLeuThr 296
 Db 1041 ATGTACAACCTCGCTGCTGATTGAAAAATAAAAAAGGCGAGCTACAGCCATTAT----- 988
 Qy 297 GluGluAspValAspTyr---AspAsnProGlyLeuSerCysGlyPheHisAspAsp 315
 Db 987 -----TCCGATTACAGCAGCGGTGCGGATTTACAACTGGCA----- 952
 Qy 316 AspAlaHisAlaHisAlaHisAsnGlyLysProTyrPheAspLeuArg---AsnLysArg 334

Db 951 -----GCCGATGTCGCTGAACACGATCTG 928
QY 335 TyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
Db 927 CTGACGTTTCCGCTTAACCTGGAAA----- 904
QY 355 LeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePhe 374
Db 903 -----GATGACGTATACACGGGAAAGGTGGCGCCGACGCGCTTACGATCGCTAT 853
QY 375 AsnAsnGlnThrGlnAlaAlaGlyLeuLeuArgHisGlnProIleGlyArgLeuLys 394
Db 852 GAAGATGCTACCTGCTGCTGCCAGTGAATATCAATGAGCTGCTGCCATATGTCGAT 793
QY 395 GlySerTrpGlyValGlnTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGlu 414
Db 792 GTTGTGGCTGAATACGATGATGCTGCGCGCATAGC-----GTAGAA 751
QY 415 AlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyVal 434
Db 750 GCGAAACACACGAGAAA-----GATGGCAGCATCACCATATATGAC----- 709
QY 435 GluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLys 454
Db 708 -----GACAC----- 703
QY 455 AlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeu 474
Db 703 ----- 703
QY 475 ProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyr 494
Db 702 -----AATCAGTCAGCTTTAACTGGCAGGTGATGGGAAATACCAC 661
QY 495 PheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThr 514
Db 660 TTGTCAATGAACACACGCTGGCGCTTCGTACTATGACGCAACAGCTTTCGACGCTG 601
QY 515 GlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys 534
Db 600 AAAGAACGCTATACCATCCAA-----CCTGCGTATAACACAGATAGCGATTGTTAACCCG 544
QY 535 HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu-----GlyTyrGluGlyAsp 553
Db 543 CAGCTCAACCGGAACCGCTCGCGGGTGATTTAACTGGGAATGGCTTCACGAC 484
QY 554 ArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln 573
Db 483 GACTGGGGATTTCAGGTCAGCGTTTACTATAACCGGTGAGTGATGCCATCCTCTCGCAC 424
QY 574 ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuVal 593
Db 423 -----AATATCGATCGCATACC-----ATTCAA 400
QY 594 ArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysPro 613
Db 399 AATCAGAACAGCGCAGCGGTGATTTACAGCGGTCTGATGCCGTTATTAGGGGAAATC 340
QY 614 ThrProArgTyrArgIleGlyValSer-----GlyAspTyrValArg--- 627
Db 339 AGCAATATATCGATGTAGGATTGATGAGCTACGCCCTGATCCACGTCAGCCAAACGTAAA 280
QY 628 -----GlyArgLeuLysAsnLeuPro----- 634
Db 279 GACATCGGCAAGATAACCGATCTGCCAACACGACAAATGACCGATGACTCTCAAA 220
QY 635 -----SerLeuProGlyArgGluAspAlaTyrGlyAsn 645
Db 219 CCGTGGGAGCGGTAAACGCTCTCGGAAGAGCGCGTTCCTCCACGTCAGTAAAC 160
QY 646 ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeu--- 664
Db 159 -----AGTACGCGTTTACAAAAAGCCCGCGTTTTCGCGGTGACCCACATTCGA 112

QY 665 -----GlyPheHisLeuLysAlaSerLeuThrAspArgIleAsp 677
Db 111 GCCGATTACACCTTAGTCTAGCTTCAGCGTCAATGCGTCAATAACCTGTTTGTAT 52
QY 678 AlaAsnLeuAspTyrTyrArgValPheAlaGln 688
Db 51 ACCCAATATGCTACATGTAAGGGTTTGTGTAA 19

RESULT 23

US-10-085-959-109/c
; Sequence 109, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 109
; LENGTH: 76804

TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: misc feature
LOCATION: (39751)..(39751)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (49372)..(49372)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (66539)..(66539)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73693)..(73693)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73705)..(73705)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73737)..(73737)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73759)..(73759)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73761)..(73761)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73766)..(73766)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73774)..(73774)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73781)..(73781)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73798)..(73798)
OTHER INFORMATION: Unsure
NAME/KEY: misc feature
LOCATION: (73802)..(73802)
OTHER INFORMATION: Unsure
US-10-085-959-109

Alignment Scores:

Pred. No.: 3,75e-05

Length:

76804

Score:	164.50	Matches:	163
Percent Similarity:	32.78%	Conservative:	115
Best Local Similarity:	19.22%	Mismatches:	297
Query Match:	4.08%	Indels:	23
DB:	12	Gaps:	473
US-09-936-377-2 (1-758) x US-10-085-959-109 (1-76804)			
Qy	6	LeuLysProlleValValLeuSerlleLeuLeuIleAenThrProLeuLeuAlaGlnAlaHis	25
Db	57074	CTTGGTCCGCTGCTTCTTCGTGTG-----TCACCAAGCTGTCCGCCACCAAAAC	57024
Qy	26	GlutThrGlnGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArg	45
Db	57023	GATGATAATGAG-----ATCATAGTGTCTGCCAGCGCGACCAAT	56985
Qy	46	-----AlaThrSerGlyLeuLeuHisThrSerThrAlaSerAsp	58
Db	56984	CGAACTGTACGGGAGATGGCGAAACCACTCGGTGTATCGAAATTCGGAATCGAGCAG	56925
Qy	59	LysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAsp	78
Db	56924	CAGATTCAGGCGCGTAAAGAGCTGAAGAGCGACTGGCTCAGTTA-----	56880
Qy	79	GlyValProGlyIleHisAlaSer-----GlnTyrGly---GlyGlyAla	92
Db	56879	---ATCCCCGGCTTGATGTTCAGACCCAGACCGCAACCACTACGTTATGAACATGCGT	56823
Qy	93	SerAlaProValIleArgGlyGlnThrGlyArgGlnIleLysValLeuAsnHisGly	112
Db	56822	GGCGCCCGCTGTGTCTGATTGACGGTGTGCGCTCAACTCTTCAGTTCCGACAGC	56763
Qy	113	GlutThrGlyAspMetAlaAsp---PheSerProAspHisAlaIleMetValAspThrAla	131
Db	56762	CGCAACTGGACTCTGTCGATCTCTTTTAAATATCGACCAT-----	56724
Qy	132	LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn	151
Db	56723	-----ATTGAAGTGATCTCCGCG---CGACGCGCCTGTACGGTGGCGGAGT	56679
Qy	152	ValAlaGlyLeuValAspValAlaAspGlyLys---IleProGluLysMetProGluAsn	170
Db	56678	ACCGAGGGTTGATCAATCATCTGTGACCCAAAAGCCGCCGGAACCATGATGAGTTT	56619
Qy	171	GlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlnLysLeu	188
Db	56618	GAGGCTGCACAAAAGATGGCTTTAAACAGCAGTAAAGATCAAGATGAGCGCATTC	56559
Qy	189	---ThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGly	207
Db	56558	GCTGTCTCCGCGGAATGAC-----CATATCTCCGGA-----	56526
Qy	208	LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro	227
Db	56525	-----GCTCTTCC-----	56517
Qy	228	AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg	247
Db	56516	GTGGCATATCAGAAATTTGGC-----GGCTGGTTGACCGT	56481
Qy	248	PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis	267
Db	56480	AACGGCGATGCCACCTGCTTGAT---AACACCCAGACCGGCTG---CAGCACTCCAAT	56427
Qy	268	GlutYrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr	287
Db	56426	CGGCTGGACATCATCGGAACCGGTACGCTG-----AACATCATGAATCCCGGCAG	56376
Qy	288	LeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeu	307
Db	56375	CTTCAACTGATAACGCAGTACTATAAAGTCAG-----GGGACGACCAATTCAGGCTT	56322
Qy	308	SerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyr	327

Db 55388 ---ACCATCAGTGTGAAGACGACAGCGCCGCGGTATTATACGGCGTGAAGGTGCGGTGGAC 55332
QY 610 TyrPheLysProThrProArgTyrArgIleGly----- 620
Db 55331 TACCTGATCCGGATACCTGACTGGAGTACCGGTGTGNACTTCAATGTGCTGAAACCGAG 55272
QY 621 -----ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGly 638
Db 55271 TCGAAAGTGAACGCTCAATGGCAAAATATAGCGTGAAGGAA-----TCAAGTCCATCG 55218
QY 639 ArgGluAspAlaTyr----- 643
Db 55217 AAAGCGACGCTTACATTAACTAGGCGCGCGGAACCGTGGAGTCTGGGTACAGAGACC 55158
QY 644 -----GlyAsn-----ArgProPheIleAlaGlnAsp 652
Db 55157 ACTTCCTTCGACGTAGCGATGACAGGGTAAACGATATTAAATGGTTACACTACCGTGCAT 55098
QY 653 AspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeu 672
Db 55097 TTTATCAGTAGTTGGCAGCTTCGCGTGGGAACACTCAGCTTCAGCGTTGAG---AACCTC 55041
QY 673 ThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 55040 TTGACCGCT-----GACTATACCACTGTCTGGGGACGCGTGCACCTCTG 54996
QY 693 ArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeu----- 706
Db 54995 TACTAC-----AGCCCGGTTACGGCCCTGCTTCACTGTACGACTACAAAGGCGG 54945
QY 707 -----GlyAlaAsnTyr 710
Db 54944 GGCGAACCTTTGGTCTGAACCTAC 54921

RESULT 24

US-10-043-344-3
; Sequence 3, Application US/10043344
; Publication No. US20030089086A1
; GENERAL INFORMATION:
; APPLICANT: Loesmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murdin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MTS
; CURRENT APPLICATION NUMBER: US/10/043,344
; PRIOR FILING DATE: 2002-07-01
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(2100)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2117)..(4852)
US-10-043-344-3

Alignment Scores:
Pred. No.: 1,34e-06 Length: 5009
Score: 161.50 Matches: 199
Percent Similarity: 32.32% Conservative: 130
Best Local Similarity: 19.55% Mismatches: 318
Query Match: 4.00% Indels: 371
Gaps: 54

US-09-936-377-2 (1-758) x US-10-043-344-3 (1-5009)
QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 2120 ACTAAAAAACCTTATTTTCGCCCTAAGTATTATTTCTGTCTTTAAATTCATGCTATGTA 2179
QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2180 AAAGCAGAACTCAAAGTATAAAAGATACAAAAGAAAGCTATATCATCTGAAGTGGACACT 2239
QY 33 -----LeuGluThrValThrValValGly-----LysSerArgPro 44
Db 2240 CAAAGTACAGAAAGATTAGAAACTATCTCAGTCTACTGCAGAAAATAAAGAGAT 2299
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2300 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATCAAAACTAGT 2347
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2348 GAAAGTATCAGCGGAGAACAAAGTATTAAATATTTCGTGATCTAACACGCTATGATCCAGG 2407
QY 83 IleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal-----IleArgGlyGlnThr 101
Db 2408 ATTTCAGTTGTAGAACAAAGTTCGGGTGCAAGTCTCGGATATTCTATTCGTGGTATGGAC 2467
QY 102 GlyArgArgIleLysValLeu----- 108
Db 2468 AGAAATAGAGTTCGTTTATTAGTAGATGGTTTACCTCAAAGCAATCTTATGTAGTCAA 2527
QY 109 -----AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 2528 AGCCCTTTAGTTCGTCGTTCCAGGATATTCTGGCATATTCTGTCATTAATTAATGAATAT 2587
QY 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro 142
Db 2588 GAAATGTA-----AAGGCCGTGCAATTAAGCAAGGGGGG 2623
QY 143 ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal----- 156
Db 2624 AGTCTCTCTGAGTATGCTAATGAGCAGTACTAGTCTGTCTGTAACTTTCAAAGCAATCA 2683
QY 157 -----AspValAlaAspGly-----LysIleProGluLys 166
Db 2684 GCAGCCGATATCTTAGAAGGAGACAAATCATGCGGAATTCAAACTTAAATATGCTTATTCA 2743
QY 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186
Db 2744 AGCAAAATAAAGGCTTTTACCCTTCTTAGCTGTAGCAGGAAAACAAGGTGGATTGAA 2803
QY 187 LysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheVal-----Leu 203
Db 2804 -----GGCTAGCCATTTACACTCAACGAAATTCAAATGAAACCCCAAGTC 2848
QY 204 HisThrGluGlyLeu-----TyrArgLys 211
Db 2849 CATAAAGATGCATTAAGGCGGTACAAAGTTATGATCGATTAAATCGCCACACAGATAAA 2908
QY 212 SerGlyAspTyr----- 215
Db 2909 TCTTCAGGATACTTTGTGATACAAGGTGAGTGTCCAATGGTGTATGACAAGTGTGCAGCC 2968
QY 216 -----AlaValProArgTyrArgAsn 222
Db 2969 AAGCCACCTCGACTTTTCCACCAAGCGAAACCGTAAGCGTTTCAGATTATACGGGG 3028
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsn----- 235
Db 3029 GCTAACCGTATCAAAACCTAATCCAATGAAATGAAAGCCAGCTCTGGTTTTTAAGAGGA 3088
QY 236 GlyGlnHis-----ArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThr 252

Db 3089 GGGTATCATTTTCTGAACAACATTATATATTGGTGGTATTTTGAATTCACACACAAAA 3148
Qy 253 Tyr-----SerAspArgAsp--- 258
Db 3149 TTTGATATCCGTGATATGACATTTCCCGCTTATTTAGCCCAACAGAAAGCGGATGAT 3208
Qy 259 -----GlnTyrGlyLeuProAlaHisSerHisGlu--- 268
Db 3209 AGTAGTCGTTCTTTTATCCAATGACAGATCATGGTGCATATCAACATATATTGAGGATGGC 3268
Qy 269 -----TyrAspAspCysHis----- 273
Db 3269 AGAGCGGTAAATATATGCAAGTGGCTTTATTTTCATGTAACACCATAGAGAAACACCGTGTA 3328
Qy 274 -----AlaAspIleIle----- 277
Db 3329 GGTATTGAATATATTTACGAAATTAAGAACAAAGCGGCATCATTTGACAAAGCAGTGTTA 3388
Qy 278 -----TrpGlnLysSerLeuIleAsnLysArgTyrLeuGln-----Leu 290
Db 3389 AGTGCTAATCAACAAACATCATCTTGACAGTTATATGCGACATATCGCATTTGCAGTCTT 3448
Qy 291 TyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGly 310
Db 3449 TATCCT-----AATCCAAGTAAGAATTGCCGC 3475
Qy 311 -----PheHisAspAspAspAlaHisAlaHisAla 321
Db 3476 CCAACACITGATAAACCTTATTCATCTATCTGATAGAAATGTTTATAAGAAAAA 3535
Qy 322 HisGlnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
Db 3536 CATAAT-----ATGTTGCAATTG---AATTTAGAGAAAAAATTCACAAAAATTGG 3583
Qy 342 -----LysGlnProPhePro---GlyPheGluAlaLeuArgValHisLeuAsnArg 357
Db 3584 CTTACTCATCAATTGCTTCAATCTGTTTGTGATGACTTACTTCAGCGCTTCAGCAT 3643
Qy 358 AsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGln 377
Db 3644 AAGATTAT----- 3652
Qy 378 ThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrp 397
Db 3653 ---TTAACTCCGAGTGTATCGCTACGCGAGATAGTATTCCCAAGGAAACCTGGTGAAACT 3709
Qy 398 GlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerSerGluAlaValLys 417
Db 3710 GGTAAACCAAGAAATGGTTGCAATCA----- 3736
Qy 418 GlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAla 437
Db 3737 CAACCTTACTTTATACCCA---AAACAGCAGCCATAT---TTTGCAGGACAGATCATTTGT 3790
Qy 438 AsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIle 457
Db 3791 AATTATCAA-----GGTACTCTCTTAATTACAGACACTGTTAAAGTG 3832
Qy 458 ArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr----- 470
Db 3833 CGG-----TTAATTAAAGGMAAAATTTATTTTCGCGACGCAATAATATG 3880
Qy 471 ---LysGlnProLeuProAspLeuGly----- 478
Db 3881 GCATTAGGGAATAVACGTTGATTTAGGTTTTCGGTATTCGGTATTCGGTATTCCTCGTACAAAA 3940
Qy 479 AlaHisArgGlnThrAla-----ArgSerPheAlaLeuSerGlyAsnTrp 493
Db 3941 GGTAAAGTCACTATAGTGTGGTAAATTTAAAAATTTCTCTGGAATACTGGTATT 4000
Qy 494 TyrPheThrProGlnHisLysLeuSerLeuThr-----AlaSerHisGlnGluArgLeu 511
Db 4001 GTCATAAAACCAACGAAGTGTGATCTTTCTTATTCGCGCTTTCTACTGGATTAGAAAT 4060

Qy 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 4061 CCTAGTTTCTCTGAATGTAT-----GGTTGGCGGTATGGTGGCAAGATGACAGGTT 4114
Qy 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 4115 TATGTAGTAAATTTAAAGCCTGAAACATCTCGTAACCAAGAGTTTGGTCTCGCTCTAAAA 4174
Qy 552 GlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIle--- 570
Db 4175 GGGGATTTTGTATATATGAGATCAGTCATTTTAGTAATGCTTATCGAAATCTTATCGCC 4234
Qy 571 TyrAlaGlnThrLeuAsn-----AspGlyArgGlyProLysSerIleGluAspAsp 587
Db 4235 TTTCTCGAAGAACTTAGTAAAAATGGAACCTGGAAGGGCAATTATGGATATCATATAATGCA 4294
Qy 588 SerGluMetLysLeuValArgTyrAsn---GlnSerGlyAlaAspPheTyrGly----- 604
Db 4295 CAAAATGCAAAATTAGTTGGCGTAAATATACTCCACAAATTAGATTTTAATGGTTATGG 4354
Qy 605 -----AlaGluGlyGluIleTyrPheLys--- 612
Db 4355 AAACGTATTCCTACGTTGGTATGCAACATTTGCTTATACCAAGTAAAGTTAAAGAT 4414
Qy 613 -----ProThr 614
Db 4415 CAAAAATCAATGCTGTTTAGCCTCCGTAAGCAGTATTTATTATTGATGCCATTCAGGCC 4474
Qy 615 ProArgTyrArgIleGlyValSerGlyAsp----- 624
Db 4475 AGCGGTATATCATTTAGTTTGGTATGATCATCCAGTAATATCTCTGGGGAATTAATACA 4534
Qy 625 ---TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 4535 ATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTCAGGAAACGT---GCATTA 4591
Qy 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 4592 GGTAAACAAT-----TCAAGGATGTAATAATCAACAGAAAACTTACTCGGCA--- 4639
Qy 664 LeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyr 683
Db 4640 -----TGGCATATC-----TTAGATGTATCG 4660
Qy 684 ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMet 703
Db 4661 GGTATTATACATGCGCAATAAA-----AATATTATG 4690
Qy 704 LeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLeuTrpAsnTrpTyrVal 723
Db 4691 CTTTCGATTAGGATATATAATTTTATTCACATATCGCTATGTTTACTTGGGAACGGTGCCT 4750
Qy 724 Lys---AlaAspAsnLeuLeuAsn-----GlnSerValTyrAlaHisSerSerPheLeu 740
Db 4751 CAAACAGCAACAGGTGGCGTCAATCAACATCAAAATGTTGGTAGCTATCTACTCGTACGCA 4810
Qy 741 SerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 4811 GCATCA-----GGACGAACATATACCTTAACATTTAGAAATGAAATTC 4852

RESULT 25

US-10-043-344-2
; Sequence 2, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loonsmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Mordin, Andrew D.
; APPLICANT: Klein, Michel H.

Qy	5	ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu	20
Db	2168	ACTAAAAAAACCCATTATTTTCGCCCTAAGATTATTTCTGTCTTTTAAATTCATGCTATGTA	2227
Qy	21	LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly-----	32
Db	2228	AAAGCAGAAGCTCAAAGTATAAAAGATACAAAGAGCTATATCATCTGAAGTGGACACT	2287
Qy	33	-----LeuGluThrValThrValValGly-----LysSerArgPro	44
Db	2288	CAAAGTACAGAGATTTCAGAAATTAGAAACTATCTCAGTCACTGCAGAAAAAATAAGAGAT	2347
Qy	45	ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser	62
Db	2348	CGTAAAGATATTAAGAGTA-----ACTGGACTTCGGCAAAATATCAAAACTAGT	2395
Qy	63	GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly	82
Db	2396	GAAAGTATCAGCCGAGAACCAAGTATAAATATTCGTGATCTAACACGCTATGATCCAGGG	2455
Qy	83	IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr	101
Db	2456	ATTTCAGTTGTAGAACCAAGTCGCGGTGCAAGTCTTGATATTTCTATTCGTGTATGGAC	2515
Qy	102	GlyArgArgIleLysValLeu-----	108
Db	2516	AGAAATAGAGTTCGTTATTATAGTAGTGGTTTACCTCAACGCAATCTTATGTAGTCCAA	2575
Qy	109	-----AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro	122
Db	2576	AGCCCTTTAGTTCGTTCCAGGATATTCGGCAATTCGTCGAATTAATGAATTTGAATAT	2635
Qy	123	AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro	142
Db	2636	GAAAAATGTA-----AAGGCCGTGCAAAATAACCAAGGGGGGG	2671
Qy	143	ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal-----	156
Db	2672	AGTCTCTCAGATATGGTAATGGAGCACTAGCTGGTCTGTACACTTCAACGCAANTCA	2731
Qy	157	-----AspValAlaAspGly-----LysIleProGluLys	166
Db	2732	GCACCGGATATCTTAGAAGAGCAAAATCATGGGGAATTCAAACTAAATATGCTTATCA	2791

US-09-936-377-2 (1-758) x US-10-114-170-119 (1-3275)

```
QY 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys-----ArgLeu 226
DB 103 TATCGCATCAACGGTAGTTATTCTATCATCAGGCAATCGTGATACCGCGGATGGAAGTCTG 162
QY 227 ProAspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLys 246
DB 163 CCGAATACC-----AACTATCGTAAACATATGTCAGGGTGTATGGTGGTATTAACTCC 216
QY 247 ArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSer 266
DB 217 GGAACCATCGT-----TTTGGCCTCTCGTTGATCGCTACAGACTCCGACACAACCT 270
QY 267 HisGluTyrAspAspCysHisAlaLeuPheLeuTrpGlnLysSerLeuLeuAsnLysArg 286
DB 271 TAC--TATGAGGATCCAGACGGAATTTAT-----GAGGCA 303
QY 287 TyrLeuGlnLeuTyrProHisLeuLeuThrGluLys-----AspValAsp 301
DB 304 TTTAGTGTCAAAATACCTAACTTGAACGAGAGAAGTTGGGTATTTCTATGACACAGAC 363
QY 302 TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAla 321
DB 364 GTGGACGGTGACTATCTAAAA-----AAAATTCTATTCGACGGC 402
QY 322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
DB 403 TATGAGCAGACATCCAGCGCAATTTGCCAAC-----GAATAAAAACGACACAG 453
QY 342 LysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
DB 454 CTGTGTTCCAGTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 507
QY 362 HisAspGluLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGln 379
DB 508 CATGATAAGCAATACACTCAGCGGTGCATATGCAGAGTCACTTTTCGCTGCTGCTAAT 567
QY 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
DB 568 AATGAACCTTTGTTACC-----GGTGCA 588
QY 400 GlnTyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
DB 589 CAGTACAAACAAGATAGGTGACCAAGGTCCGGTGGCATGACCTCAAGCAAAATCTCTG 648
QY 416 ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGlu 435
DB 649 ACCGGC-----TTCATTAATAAGAAACACGAACTCGCTCTTATTTATGATCAGAG 699
QY 436 GlnAla-----AsnTrp-----AspAsnPheThrLeu 444
DB 700 CAAGTACAGTCTCACTATTCGCACAAATGACTGGCAATTCGCCGATCACTGGACATGG 759
QY 445 GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
DB 760 ACAATGGGAGTTGCG-----CAATACTGCTTCTTCAAGTTGACGGTGTGT----- 807
QY 465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
DB 808 GACGGAGTATCATATACCCGAGCATATTATAAGCATACCTCTCTTCCAGAGAGTCTGCG 867
QY 485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
DB 868 AGTGATCAGAAATGTGTAACATCTACAGCTCGCTATTTCCAGTTTCAGTTTCGATTAACCTGGAG 927
QY 503 LeuThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHis 520
DB 928 TTTACGGCTGCGTTCGGCAGAGCTAGTATTTCCACACTCTCCCGAGCTTTTATGCGAG 987
QY 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
DB 988 ACATCTCGCGGGCGGACGTGTCATACATAC-----GGAATCTCTGATCTTAAGGCTGAACAC 1041
```

```
QY 541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAla 560
DB 1042 TCCAAATAACTTTGAATTAGTGCACGATATAATGTAATCAGTGGCTGATGACACGCCA 1101
QY 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
DB 1102 GTTTACTACTCAGAGCTAAAGATTATATT-----GCAAGTCTGATCTGTGATGGCAGTATA 1158
QY 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr-----AsnGlnSer 598
DB 1159 GTTTGCAATGGTAACACCAACTCTCCCGTAGTACTACTATTATTATGACAAATATTGAT 1218
QY 599 GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArg 618
DB 1219 CGGGCAAAACATGGGGACTGGAA-----ATAAGCGCGGAATATAAT 1260
QY 619 IleGlyValSerGlyAspTyrValArgGlyArgLeu----- 630
DB 1261 GGCTGGTTTTCTCGCATATATCAGTGGCAATTTAATTCGTGCGCAATATGAAACTTCA 1320
QY 631 -----LysAsnLeuProSerLeuProGlyArg----- 639
DB 1321 ACATTAAAAACAACATAATACAGTGAACCCAGCGATAAACGGACGTATAGGGCTGAAACAT 1380
QY 640 -----GluAspAlaTyrGlyAsnArgPropheIle 649
DB 1381 ACTCTGTGATGGTCCAGCCCAACATACTCTGATGTTTATTATTCGTGCTGCTCTAGT 1440
QY 650 AlaGlnAspAspGlnAsnAla-----ProArgValProAla-----AlaArgLeuGlyPhe 666
DB 1441 GCAAAAGATGACAGTAACTACCGTACCGGAAACAATGTTCCGGCTGGGCCACTCTCACTTT 1500
QY 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal--- 685
DB 1501 GCAGTAAATACAGAAATTC-----GGTAACGAGGATCAGTACCGGATTAAAC 1545
QY 686 PheAlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly 700
DB 1546 CTGGCCTCAATAACCTGACAGCAACACGCTACCGTACAGCACATGAAACTATTCTCTGCA 1605
QY 701 -HisHisMetLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyr 716
DB 1606 GCAGGTTTTTATGACGCTATAGTGTGTTGATGGAATTTCTGATGACAAAAACAATCGTTA 1665
QY 716 rGly 717
DB 1666 TGA 1669
```

RESULT 27

US-10-238-075-719

; Sequence 719, Application US/10238075

; Publication No. US20030148324A1

; GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D- A- and which are isola

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238,075

; PRIOR FILING DATE: 2002-09-10

; PRIOR APPLICATION NUMBER: 0003145

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 719

; LENGTH: 5585

; TYPE: DNA

; ORGANISM: Escherichia coli

US-10-238-075-719

Alignment Scores:

Pred. No.: 2,03e-06 Length: 5585

Score: 160.50 Matches: 129

ORGANISM: Escherichia coli
US-10-085-959-58

Alignment Scores:

Pred. No.: 3,646-06 Length: 8249
Score: 160.50 Matches: 129
Percent Similarity: 36.25% Conservatives: 82
Best Local Similarity: 22.16% Mismatches: 239
Query Match: 3.98% Indels: 132
DB: 12 Gaps: 33

US-09-936-377-2 (1-758) x US-10-085-959-58 (1-8249)

```

Qy 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys-----ArgLeu 226
Db 116 TATGCATCAACGGTAGTTATCTGTATCAGGCAATCGTATACCGCGGATGGACGCTG 175
Qy 227 ProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLys 246
Db 176 CCGAATACC-----AACTATCGTAACAATAGTCAGGGGTATGTTGGGTATTAACATCC 229
Qy 247 ArgPheTyrArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSer 266
Db 230 GGAACCATCGT-----TTTGGCTCTCGCTTACGTACAGACTCGCGAAGCAAACT 283
Qy 267 HisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArg 286
Db 284 TAC---TATGAGGATCCAGACGGAAGCTAT-----GAGGCA 316
Qy 287 TyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu-----AspValAsp 301
Db 317 TTTAGTGTCAAAATACCTAACTTGAACGAGAGAAAGTTGGGTATTTCTATGACACAGAC 376
Qy 302 TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAla 321
Db 377 GTGACGCGTGACTATCAAAA-----AAAATTCATTTCCGACGCG 415
Qy 322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
Db 416 TATGACGACAGACCATCCAGCGCAATTTGCCAAC-----GAAGTAAAAACACACAGAC 466
Qy 342 LysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
Db 467 CCTGTTCCAGTCGATGATTCAGGCTCTGACGCTGCTAT---AACAGACTGAC---ACC 520
Qy 362 HisAspGluLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGln 379
Db 521 CATGATAAGCAATACACTCAGCGCGTCAATTGACAGAGTCACCTTTTCGCTGCTGCTAAT 580
Qy 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
Db 581 AATGAACTTGTACC-----AsnTrp-----GGTGCA 601
Qy 400 GlnTyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
Db 602 CAGTACAAACAGACAGGTCAGCAAGGTCGGTGCATGACCTCAAGCAATCTCTG 661
Qy 416 ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu 435
Db 662 ACCGCGC-----TTCAATTAATAAGGAAACACGAACCTCGCTCTATATGAGTCAGAG 712
Qy 436 GlnAla-----AsnTrp-----AspAsnPheThrLeu 444
Db 713 CAAGTACAGCTCTACTATTGCAAAATAGTCGCGATTCGCGCATGACTCGGACATGG 772
Qy 445 GluGlyGlyValArgValGlnLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
Db 773 ACAATGGAGTTGCG-----CAATAGTGGCTTTCTTCAAGATTGACGCGTGGT----- 820
Qy 465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
Db 821 GACCGAGTATCATATACCGGACGATTATTAAGCGATACCTCTCTTGGCAGAGAGTCTGCG 880

```

```

Qy 485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
Db 881 AGTGATCAGAAATGGTAACATCTACAAAGCTCGCTATTACAGGTTTCGATAACTTGAG 940
Qy 503 LeuThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHis 520
Db 941 TTACGCGCTGGTTCGCGAAGGCTAGCTATTTCCACACTCTCCAGCTTTTATGCGAG 1000
Qy 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
Db 1001 ACATCTCGGCGCGCAGTGTCAATAC-----GGAATCTCTGATCTTAAGCGCTGAAC 1054
Qy 541 SerAsnAsnIleGlnLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla 560
Db 1055 TCCAAATACCTTGAATAGGTGACGATATAATGTAATCAGTGTGCTGATGACAGCGCA 1114
Qy 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
Db 1115 GTTTACTACTCAGAAAGCTAAAGATTATATT---GCAAGTCTGATCTGTGATGGCAGTATA 1171
Qy 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr-----AsnGlnSer 598
Db 1172 GTTTGCAATGTAAACACCAACTCTCCGCTAGTAGTACTATTATTATGACAATATTGAT 1231
Qy 599 GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArg 618
Db 1232 CGGCAAAACCATGGGACTGGA-----ATAAGCGGGAATATAAT 1273
Qy 619 IleGlyValSerGlyAspTyrValArgGlyArgLeu----- 630
Db 1274 GGCTGGGTTTCTCGCCATATATCAGTGGCAATTTAATTCGTCGCAATATGAAACTTCA 1333
Qy 631 -----LysAsnLeuProSerLeuProGlyArg----- 639
Db 1334 ACATTAAAAACAATAACAGAGAACCCAGCGATAAACCGGACGTATAGGCTGAAACAT 1393
Qy 640 -----GluAspAlaTyrGlyAsnArgProPheIle 649
Db 1394 ACTCTGTGATGGGTGAGCCCAACATATCTCTGATGTTTTTATTCGTGCTGCTCTAGT 1453
Qy 650 AlaGlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPhe 666
Db 1454 GCAAAAGATGACAGTAACGTTACCGAACAATGTTCCGGGTGGCCACCTCTCAACTTT 1513
Qy 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal--- 685
Db 1514 GCAGTAATAACAGAAATTC-----GGTAAACGAGGATCAGTACCGGATTAAC 1558
Qy 686 PheAlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly 700
Db 1559 CTAGACTCAATAACCTGACAGAACCGTACCGATGACGATGAACTATTCCTGCA 1618
Qy 701 -HisHisMetLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyr 716
Db 1619 GCAGGTTTAAATGCGACTATAGGTTTCTGTATGGAATTTCTGATGACAAAAACAATCGTTA 1678
Qy 716 rGly 717
Db 1679 TGA 1682

```

RESULT 29

US-10-329-960-1/c

Sequence 1, Application US/10329960

Publication No. US20030099277A1

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Pr

TITLE OF INVENTION: Thereof, and Uses Thereof

FILE REFERENCE: PBI86PI

CURRENT APPLICATION NUMBER: US/10/329,960

CURRENT FILING DATE: 2003-01-02

PRIOR APPLICATION NUMBER: US 09/643,990

PRIOR FILING DATE: 2000-08-23

```

; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c

```



```
QY 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerG1 589
Db 48480 GATTGCTCTTATTCCTAATATGAT-----AACACCAATAGCTATGTAACACAGCGGAAA 48427
QY 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyL1 609
Db 48426 GCGCCGGTTG-----CACCGTGTGGAATTTGCCGGCACA-----48393
QY 609 eTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyAr 629
Db 48392 -----TTGCCGCTGTGTCAGAGATGTCACGCTGCTCACTGAATTACCTGGACCG 48340
QY 629 gLeuLysAsnLeuProSerLeuProGlyArgGlyAspAlaTyrGlyAsnArgProPheI1 649
Db 48339 A-----ACGTGAACACACGTGATGGT-----48321
QY 649 eAlaGlnAspAspGlnAsnAlaPro-----ArgValProAlaAlaArgLeuGlyPheH1 667
Db 48320 -----GATAACAAAGGTGCGCGCTGAGTTATACCCCTGAACACATGTTGATGCGAA 48268
QY 667 sLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheA1 687
Db 48267 ACTGAAGTGCAGATCACCGAAGAGGTGGCATCATGCTGGGT-----48225
QY 687 aGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuG1 707
Db 48224 -----GCCCGTTATCGCGGGAACACCA-----CGTTTCAC 48193
QY 707 yAlaAsnTyrArg-----ArgAsnThrArgTyrGlyGluThrPasnTrpTy 722
Db 48192 CCAGAATTATTCGTCACTAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGAGAATA 48133
QY 722 rValLysAla-----725
Db 48132 CCTGAAGCCTGGACGGTGGTGGATCGAGGTCTCTGTGGAAGATGACGGATGCCCTGAC 48073
QY 726 -----AspLeuLeuLeuAsnGln-----SerValTyrAl 735
Db 48072 GCTGAATGCTCGGTGTAATCACTGCTCAACAGGATTACAGTACGCTGTACAG 48013
QY 735 a-----HisSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrG1 752
Db 48012 TGCCGGTAAGAGTACGCTGTATGCGGTGATCTTCCAGACGGGATCATCAACACAGG 47953
QY 752 yGlyVal 754
Db 47952 ATATGTG 47946
```

RESULT 32

US-10-098-808-1

; Sequence 1, Application US/10098808

; Publication No. US20030007981A1

; GENERAL INFORMATION:

; APPLICANT: Potter, Andrew A.

; APPLICANT: Rioux, Clement

; APPLICANT: Schryvers, Anthony B.

; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS

; TITLE OF INVENTION: TRANSFERIN-BINDING PROTEINS

; FILE REFERENCE: 9000-0049.20

; CURRENT APPLICATION NUMBER: US/10/098,808

; CURRENT FILING DATE: 2002-03-13

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/405,728

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-24

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/267,749

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-10

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 6376

; TYPE: DNA

; ORGANISM: Haemophilus somnus

US-10-098-808-1

```
Alignment Scores: 7.7e-06 Length: 6376
Pred. No.: 156.00 Matches: 195
Score: 30.43% Conservative: 120
Percent Similarity: 18.84% Mismatches: 356
Best Local Similarity: 3.87% Indels: 354
Query Match: 14 Gaps: 50
DB:

US-09-936-377-2 (1-758) x US-10-098-808-1 (1-6376)
QY 6 LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeu-----21
Db 2915 CTTAAGCTGATACATTTGGCTGTCAGCAGCATTTTTTACCTTTTACTGAGCGGTTGCC 2974
QY 22 -----AlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrVal 38
Db 2975 GATACTGAATCACCGAGTAGCAATACAGAAGCAGTGTCTGGAGTTAGAACTATCCAGGTG 3034
QY 39 ValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAsp 58
Db 3035 CAAGCCAAA-----CACGAGATCAGCAGCATGAC 3064
QY 59 LysIleIleSerGly-----AspThrLeuArgGlnLys 69
Db 3065 AATGAAGTCAACCGTTTGGGTAAGGTGTCAAAAGCAGTGAAGACATTGATAAAGAACTG 3124
QY 70 AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGly 89
Db 3125 ATTTGAATATTCCGATTTGACCCCGTTATGATCCCGTATTTTCGGTGGAGCAGCGGA 3184
QY 90 GlyGlyAla---SerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValleu 108
Db 3185 CGTGGTCAACAGTCAGGCTATGCAATCGTGGTGTTCACAGAAACCGCGTGTATGTG 3244
QY 109 ---AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAsp-----His 124
Db 3245 GTGACGCGCTGGACAGCGCAGTCTCTATCTTACCTTGAATCCGATGCCAACGGCGG 3304
QY 125 AlaIle---MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db 3305 GCGATTAAATGAATGATGAGATATTAAATCAATTGAATTTGAGCAAGGGGTCCAGT 3364
QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db 3365 TCGCAGAAATACGAGTACGCGTCTTGGCGCGTGGTGGTGGTTCGTACCAAGAAAGCT 3424
QY 164 ProGlyLysMetProGluAsn-----GlyValSerGlyGluLeuGlyLeuArgLeu 180
Db 3425 GATGATGTGATTAAAGAGGGGCAAACTGGGGCTTGAACAGTAAACCGCTTACAGCAGC 3484
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGly-----196
Db 3485 AAAAACACCCAGTATTTACCCAAATCCGTTGCCGCTGCTTCCGTTCCGCGCGGTTTTCAGAGT 3544
QY 197 -----LeuGlyLysAsnPheValLeuHis-----204
Db 3545 TTGCGGATTTTATCCCATCGTAAAGGTAAAGAAACCCCGCGTCATCTCTGCTCCCAAGAA 3604
QY 205 -----ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaVal 217
Db 3605 ATACACATACCTACCAACCATTTGGAAGGATTTTATATCGG-----TATGAGGTT 3655
QY 218 ProArgTyrArgAsnLeuLys-----224
Db 3656 GACCAAAACCCGAAACGGAAGCCTGTTCTGGCAATGCGTATTTATATCTTGTCCCGATGAA 3715
QY 225 -----ArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaVal---241
Db 3716 TGCTCTAATCTAAGTATCCGAGTGTGCTCATGCCAAGGCCCAAGCAAGATAGGTGGGT 3775
QY 242 -----LeuGlyTyr-----ArgLysArg 247
```

3776 G C C C G G A G A C A A T C C T A A T T G G A C G C C C G A A G A G C A G C A C A G C G C T C T A A A A T G C G C 3835
Db
3777 P h e T y r A r g A r g T h r T y r S e r : : : : : A s p A r g A r g A s p G l n T y r 260
Qy
3836 T A T C C G A C A C G T A C C G C C T C T G C C A A G A T T A T A C G G T C C T G A C C G C A T C A G C C C T A A T 3895
Db
261 G l y L e u P r o A l a H i s S e r H i s G l u T y r : : : : : 269
Qy
3896 C C G A T G G A C T A C C A A G T C A C T C T T C T C T G T G A A A G G T G T T A C C G C T T G T C G C C T A A C 3955
Db
269 : : : : : 269
Qy
3956 C A T T A T G T C G G C G G T G T T G A A C A T A C A G A C A G C G G T T A C G A T A T C C G T G A T A G A C G 4015
Db
270 : : : : : A s p A s p C y s H i s A l a A s p I l e T r p G l n L y s S e r 281
Qy
4016 C A A C G G G C G T A T T A C A G A A G A G A T A T C T G C C A C A G C G A T C C A G T T G C C A A C G T T G 4075
Db
282 L e u I l e A s n L y s A r g T y r L e u G l n L e u T y r P r o H i s L e u L e u T h r G l u G l u A s p V a l A s p 301
Qy
4076 G A T A A A A A T G A G A C G G A A A G G T A A T T C G G T A T C A C G T T G A C T G A T A T C C T T T T G G A C 4135
Db
302 : : : : : T y r A s p A s n P r o G l y L e u S e r C y s G l y P h e H i s A s p A s p : : : : : 314
Qy
4136 G G T T G G T A T A T G A T C C G C A A T C A A G C T C G T G C G T G C G G T A C G G A C G G G T A A A T T 4195
Db
315 : : : : : A s p A s p A l a H i s A l a H i s A s n G l y : : : : : 324
Qy
4196 T T T A A T G A A C G C C A T A C G A A A A T C G C T C G G T A T C T T T A C C G T A T G A G A T C C C G A T 4255
Db
325 : : : : : L y s P r o T r p : : : : : I l e A s p L e u A r g A s n L y s A r g T y r G l u 336
Qy
4256 A A A A A T C T T G G C A G A T A G C T T G A C C T T G A G T A T T A C C G C C A A G A T C T C A A A C T G T C G 4315
Db
337 L e u A r g A l a G l u T r p L y s : : : : : G l n P r o P h e P r o G l y P h e G l u A l a L e u A r g V a l H i s 354
Qy
4316 A G C G T A T C C A T T G G A C G T A T G C A C C G A T A T C C T C A T G T G G C A C G T T G C C G T G C C A G C 4375
Db
355 L e u A s n : : : : : A r g A s n A s p T y r H i s A s p G l u L y s 365
Qy
4376 T T G G A C A A C C T T G G T C T A A T T A C C G T A C C G A A A A A C G A T T A T C A A G A C G A C T C A A T 4435
Db
366 A l a G l y A s p : : : : : A l a V a l G l u A s n P h e P h e A s n G l n T h r G l n A s n A l a A r g I l e G l u 384
Qy
4436 C T G G C A C A A T T C A A T T G G G A A A A A C T T T A A T C T G G G C T T T A C C A C G C A T A A G T G A A T 4495
Db
385 L e u A r g H i s G l n P r o I l e G l y A r g L e u L y s G l y S e r T r p G l y V a l G l n T y r L e u G l y G l n 404
Qy
4496 A T C : : : : : G C C G C G G C T T T G G T A C A 4516
Db
405 L y s S e r S e r A l a L e u : : : : : S e r A l a T h r S e r G l u A l a V a l L y s G l n P r o M e t L e u 421
Qy
4517 C A T C G C T C C A C C T T A C A C A T G G C G A C T T A T G T G A A T A T G T C A C C T T G C C A C C G T A T 4576
Db
422 L e u A s p A s n L y s V a l G l n H i s T y r S e r P h e G l y V a l G l u G l n A l a A s n T r p A s n 441
Qy
4577 A C A G A G A A A A A G T G : : : : : T A T G C G A A G A T A T A A G T C A A A C A A A A T 4621
Db
442 P h e T h r L e u G l u G l y G l y V a l A r g V a l G l u L y s G l n L y s A l a S e r I l e A r g T y r A s p L y s 461
Qy
4622 C C G A C A G C A G A A : : : : : G A A A A A G A G A A : : : : : T T A C A T A C G C C A A T 4660
Db
462 A l a L e u I l e A s p A r g G l u A s n T y r T y r L y s G l n P r o : : : : : L e u P r o A s p L e u : : : : : 477
Qy
4661 G G T T C T T A T G A C A A C C T C G G T A T A T A G A C G T A A A A A C A C G C C G G A T T A A A A A C T G T C 4720
Db
478 : : : : : G l y A l a H i s A r g G l n : : : : : T h r A l a 484
Qy
4721 A A T G G T G C A A T G A C A C A G C A G G C G A T A A C C G T G A C T G C T C C C A C G T G T G A T A C G G C 4780
Db
485 A r g S e r P h e A l a L e u S e r : : : : : 490
Qy
4781 A G A C A G T A T A C C T T G C C T T G G T A A C C A T A T T G C C T T G G T G A A T G G C A G A C T T G G G G 4840
Db

490 : : : : : 490
Qy
4841 T T G G C G T C G G T A C G A C A A C C A T A C T T C C G C T G A A T G A C C C G T G G A C C C A A A G T G G C 4900
Db
491 : : : : : G l y A s n T r p T y r P h e : : : : : T h r P r o G l n H i s L y s 500
Qy
4901 A A C T A C C A C A C T G T G C T G G A A T G C G G G C G T G A G C C T C A A A C C A C C C G C C A C T T T G T C 4960
Db
501 L e u S e r L e u T h r A l a S e r H i s G l n G l u A r g L e u P r o S e r T h r G l n G l u L e u T y r : : : : : 518
Qy
4961 G T G T C T A C C G T G T C C A G C G G T T T C C G T C C C G C T T T T A T G A G C T G T A C G C G T G 5020
Db
519 : : : : : A l a H i s G l y L y s : : : : : H i s V a l A l a T h r A s n T h r P h e G l u V a l G l y A s n L y s 534
Qy
5021 C G T A C G G G C T T C T G T A A A G A C A A T C C A C A C A A A A G A G T T C T T G A C C C G T A A A 5080
Db
535 H i s L e u A s n L y s G l u A r g S e r A s n A s n I l e G l u L e u A l a L e u G l y T y r G l u G l y A s p A r g 554
Qy
5081 C C G T T G A A A A G C G A A A A G C C T T T A C C A A A A T T G G T T G C C G T C A G G C G A T T T 5140
Db
555 T r p G l n T y r A s n L e u A l a L e u T y r A r g A s n A r g P h e G l y A s n T y r I l e T y r A l a G l n T h r 574
Qy
5141 G G T G T G A T A G A C A G A T T T C T T C C A A A C A C A C T A T A A A A C C T G C T : : : : : 5188
Db
575 L e u A s n A s p G l y A r g G l y P r o L y s S e r I l e G l u A s p S e r G l u M e t L y s L e u V a l A r g 594
Qy
5189 : : : : : G C C C G T C A G A T A A A T A T G T C G A G G A T T G G T T A T G T A : : : : : 5227
Db
595 T y r A s n G l n S e r G l y A l a A s p P h e T y r G l y A l a G l u : : : : : 606
Qy
5228 : : : : : A C C G A T T T T T A C A C C C A C C A G A T C T C A A C T C A A C G G T A T C A A T 5272
Db
607 : : : : : G l y G l u I l e T y r P h e L y s : : : : : P r o T h r P r o A r g T y r : : : : : 617
Qy
5273 A T C T T G G G T A G A T C T A C T G G A A G C A T C A G C A T A G G C T G C C T G A A G C T T G A T T C C 5332
Db
618 : : : : : A r g I l e G l y V a l S e r G l y A s p T y r V a l A r g G l y A r g L e u L y s A s n 632
Qy
5333 A C A C T T G C T T A C A C C G T A T C A A T A T C A A A G C A C G C A A A T T G C A C A C A A T T T A C C A A T 5392
Db
633 L e u : : : : : P r o S e r L e u : : : : : P r o G l y A r g G l u A s p A l a T y r G l y A s n A r g 646
Qy
5393 G T G T C T G A C C G C A C T T G G A A G C A G C A C C G G A C G C A T T A T T G C A : : : : : 5440
Db
647 P r o P h e I l e A l a G l n A s p A s p G l n A s n A l a P r o A r g V a l P r o A l a A l a A r g L e u G l y P h e 666
Qy
5441 : : : : : A G T A T C G C T A T G A T G A C : : : : : C C T G A G G G C A G A T G G G C C T T 5479
Db
667 H i s L e u : : : : : L y s A l a S e r L e u T h r A s p A r g I l e A s p A l a A s n L e u A s p 681
Qy
5480 A A T T T A A G C G C A C C T A C T C T C A A G C C A A A C A C A C G T A C G A A G T G G T C G G C G A A : : : : : 5533
Db
682 T y r T y r A r g V a l P h e A l a G l n A s n : : : : : L y s L e u A l a A r g T y r G l u T h r A r g T h r 698
Qy
5534 : : : : : A A A G T T C G C A G G T G C A G C A T T A A C G R C A T C A C A G C A A A C G C A C T 5587
Db
699 P r o G l y H i s M e t L e u A s n L e u G l y A l a A s n T y r A r g : : : : : 711
Qy
5588 C G T C T T G T A T A T T T A T G A T T T G A C G C A T A C A C T T G C A A A A A A A A T T C A C G T T G 5647
Db
712 : : : : : A r g A s n T h r A r g T y r G l y G l u T r p : : : : : A s n T y r T y r V a l 723
Qy
5648 A G A C C G G T A T C T A T A A T T T A A C C A A T C G T A A A T A T A G C A T G G A A A G T G T G C G T C A G 5707
Db
724 L y s A l a A s p A s n L e u A s n G l n S e r V a l T y r A l a H i s S e r S e r P h e L e u S e r A s p T h r 743
Qy
5708 T C C G C T G C C A A T G C G T C A T C A A G A C C T A G T A C A C G T T G C G C A C G T T T T G C C : : : : : 5761
Db
744 P r o G l n M e t G l y A r g S e r P h e T h r G l y V a l A s n V a l L y s P h e 758
Qy
5762 : : : : : G C A C G G G C A G A A A C T T T A C C G T G A T G A T G A A A T G A A G T T T 5803
Db

RESULT 33

US-10-238-075-1076
 ; Sequence 1076, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
 ; FILE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238, 075
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; PRIOR FILING DATE: 2000-03-10
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1076
 ; LENGTH: 2046
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; US-10-238-075-1076

Alignment Scores:
 Pred. No.: 1,59e-06 Length: 2046
 Score: 155.50 Matches: 158
 Percent Similarity: 31.8% Conservative: 108
 Best Local Similarity: 18.94% Mismatches: 248
 Query Match: 3.85% Indels: 321
 DB: 12 Gaps: 43

US-09-936-377-2 (1-758) x US-10-238-075-1076 (1-2046)

QY 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
 DB 130 CTCAGAAATGCCCGCCAGTGTCTCAGTCATTACTTACAGCACTGCAAAACAACTGCAAAACAAACCG 189
 QY 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGly 90
 DB 190 GTTTC-----GATCTGTCAGTCAGTAAGAAGTGTGAAGGGATTAGTATCACTGTT 243
 QY 91 GlyAlaSerAlaPro-----ValIleArgGlyGlnThrGlyArgArgIleLysValLeu 108
 DB 244 GGGATGAATAACCGGATATCATGATACGTGGTCTAAGTGGCGATTACACGCTGATTCTG 303
 QY 109 -----AsnHisGlyGlnThrGlyAspMet 117
 DB 304 GTCGATGGACGCTCAGACGCTCGGGAATCCAGAACCAACGACGCGCGGTTTTGAA 363
 QY 118 AlaAspPheSerProAsp---HisAlaIleMetValAspThrAlaLeuSerGlnGlnVal 136
 DB 364 GCGGATTTATCCCTCTCTGTGGAGCAATT-----GAACGCATT 402
 QY 137 GluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal 156
 DB 403 GAAGTGATCGTGGCCCTATGCTCCCTGTATGTTCTGATGCCATCGAGGGGTCATT 462
 QY 157 AspValAlaAspGlyLysIleProGluLys----- 166
 DB 463 AATATCAACCAACACAGATTAAATACCAAAACATGGGATGGCGTACTTGGACTTGGGGGG 522
 QY 167 ---MetProGluAsnGly----- 171
 DB 523 ATTATTCAGGAACATGGAAATTTGGTAATCAACCAACAAATGACTTCTATCTGTGAGGC 582
 QY 172 -----ValSerGlyGluLeuGlyLeuArgLeuSerSerGly---AsnLeuGluLysLeu 188
 DB 583 CCATTGATTAAGGATAAATCTGTCTTCACTATATGAGGAATGAATATCGCAAGAA 642
 QY 189 ThrSerGlyGlyIleAsnIleGly----- 196
 DB 643 GATAGT-----ATCTCTAGGGAACACCGGCAAAAGATTAATGAATATTAACGGCAACG 696
 QY 197 -----LeuGlyLysAsnPheVal 202

DB 697 CTCAGTTTACTCCGACTGAAGCCAGAAAGTTTGTGTTTGAATATGAAAAATAACACG 756
 QY 203 LeuHisThr-----GluGlyLeu-----TyrArgLysSerGlyAspTyrAla 216
 DB 757 GTGCATACATTAAACACACCTGGTGAGTCTCTCGATGCTGGACTATCGGGA----- 807
 QY 217 ValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGly 236
 DB 808 -----AATCTTAACAA-----CCAAACAGTAAGAGAAACGACGATATTC 849
 QY 237 GlnHisArgAlaValLeuGlyTyr----- 244
 DB 850 CGTAGTCACCTGGTAGCAGCATGGAATGCCAGGGCGAAATAGTCATCTCTGAAATTGCT 909
 QY 245 -----ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeu 262
 DB 910 GTTTATCAGGAGAAAGTTATTCTGAGGTAAATCAGGTAAAGTAAAGATAATATAAT--- 966
 QY 263 ProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGln 279
 DB 967 -----CAITGGGATCTTAATTACGAGTCAAGAAA 996
 QY 280 -----LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
 DB 997 CGGAAATAACCAACACAAATCATAGATGCAAAAGTACGCGCATTTCTGCCG----- 1047
 QY 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp 315
 DB 1048 -----GAAATGTA-----CTGACCATCGA----- 1068
 QY 316 AspAlaHisAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyr 335
 DB 1068 ----- 1068
 QY 336 GluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu 355
 DB 1068 ----- 1068
 QY 356 AsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsn 375
 DB 1069 -----GGTCAATTTTCAGCATCAGAGCTCCGTCGATGACTCAGCCACGGGTAAACAAACG 1122
 QY 376 AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly 395
 DB 1123 ACMGAACACAGCTCTGTTCAATTAACAGAAAGCTGTTTATAGAAAT----- 1173
 QY 396 SerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
 DB 1174 -----GAATATGACGACACGAGATTCTCTGCCCTGACTGGA----- 1209
 QY 416 ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu 435
 DB 1210 -----GGACTGCTCTCGATAAT-----CATGAAATCTATGGC----- 1242
 QY 436 GlnAlaAsnTyr-----AspAsnPheThrLeu 444
 DB 1243 ---AGTTACTGGATCCAGATTGTACGCTGTTTATAACCTGACCGCATATCTCAGACTC 1299
 QY 445 GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
 DB 1300 AAAGGGGGATCGCAAAAGCATTTCCGGCTCTCTCAATTCGT----- 1341
 QY 465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
 DB 1342 -----GAGGTGAGTCTGGATTTCGGCTCTCTCAATTCGT----- 1380
 QY 485 ArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGln-----HisLys 500
 DB 1381 GCCTCTATT---ATGATGGAACAGGACCTGAAACCGGAGACAGTGTAAACCGAGAG 1437
 QY 501 LeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 DB 1438 ATCGGTATTATTATAGTAATGATAGTGGTTTTTCGGCGAGCGCGACCTGTTTAACT 1497

```

QY 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys----- 534
Db 1498 GATTTTAAATAAGTTGACCACTAGCATATAGGTACAAAAGATCCAGTCACGGGTTA 1557
QY 535 -----HisLeuAsnLysGluArgSerAsnAsnLleGluLeuAlaLeu 548
Db 1558 AACACTTTTATTTATGATAATAGGTAGGAGCAAAATATCAGAGGGGTGGAGCTTGCNACT 1617
QY 549 GlyTyrGlu---GlyAspArgTTPGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
Db 1618 CAGATTCCTGTATGATAAATGCGATGATCT-----GCA 1653
QY 568 AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAsp 587
Db 1654 AACTAT-----ACATTTACTGACTCTCGT-----CGAAAAAGTATGATGACGAA 1695
QY 588 SerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGly 607
Db 1696 AGT-----CTCAATGGCAAGTCGCTGAAAGGGGAACCT--- 1728
QY 608 GluIleTyrPheLysProThrProArgTyrArgIleGly-----ValSerGlyAspTyr 625
Db 1729 -----CTGGAAGAAGACTCCAGACATGCACCAATGCAAAACCTGGAATGGATTAC 1779
QY 626 ValArgGlyArgLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn 645
Db 1780 ACTCAGGAT---ATTACATTTTATTTCATCTCTG-----AATTATACGGGAAAA 1824
QY 646 ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGly 665
Db 1825 CAATCTGGGAGCACAC-----AGAAATGGTGTAAAGTTCCCGCGTTCGTAATGGA 1878
QY 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db 1879 TTC-----ACATCTATGGATATGCTCTAAATTACCAG----- 1911
QY 686 PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsn 705
Db 1912 ATTCTGCAGACACGCTGATTAATTTTCC-----GTTCTTAAC 1950
QY 706 LeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsnTyrTyrValLysAl 725
Db 1951 GTCAC-AGACAGAAAGACGAGGATATCGATACCATGATGTAATCTGCGAGGTC----- 2004
QY 725 aAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProG1 745
Db 2005 -----GA 2006
QY 745 mMetGlyArgSerPheThrGlyValAsnValLysPhe 758
Db 2007 TGAAGGACGCGCTTATTGGCGTAAATGTAAGAGTATCCTTC 2046

RESULT 34
US-10-240-218-5
; Sequence 5, Application US/10240218
; Publication No. US20030186848A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: Goringe, Andrew Richard
; APPLICANT: Hudson, Michael John
; APPLICANT: Matheson, Mary Anne
; APPLICANT: Robinson, Andrew
; APPLICANT: West, David McKay
; TITLE OF INVENTION: Recombinant iron uptake proteins
; FILE REFERENCE: GWS/22057
; CURRENT APPLICATION NUMBER: US/10/240,218
; CURRENT FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: GB 0007433.6
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5

```

```

; LENGTH: 2748
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-240-218-5

Alignment Scores: 2,47e-06 Length: 2748
Pred. No.: 155,50 Matches: 196
Score: 32,90% Conservative: 136
Percent Similarity: 19,43% Mismatches: 333
Best Local Similarity: 3,85% Indels: 346
Query Match: 12 Gaps: 49
DB:

US-09-936-377-2 (1-758) x US-10-240-218-5 (1-2748)
QY 2 AlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuLeuLeuLeu 21
Db 3 GCAACAGCAACATTTGTCGATTCAATATTTTATGCTGTCTTTTAAATGACTCGCTGCC 62
QY 21 uAlaGlnAlaHisGluThrGlu-----GlnSerValGlyLeuGluThrValTh 37
Db 63 CGCTTATGCAAAATGTCAAGCCGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 122
QY 37 rValValGlyLysSer-----ArgProArgAlaThrSerGlyLeuLeuHisTh 53
Db 123 GGTAAAGCCAAATAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 174
QY 53 rSerThrAlaSerAspLysIlelle---SerGlyAspThrLeu---ArgGlnLysAlaVa 71
Db 175 -----GGCAAGTTGGTCAAGTCTTCCGATACGCTAAAGTAAGCAAGGTTT 221
QY 71 lAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyG1 91
Db 222 GAATATCCGAGACCTGACCCGCTTATGATCCGGGTATTCGCGTGTGCAAGGCGCGG 281
QY 91 yAlaSerAlaProVal---lleArgGlyGlnThrGlyArgArgIleLysVal----- 107
Db 282 CGCAAGTTCGGCTATTCAATACGCGCATGATAAAACCGCGTTTCTTAAACGGTGA 341
QY 108 -----LeuAsnHisHi 111
Db 342 CGGCGTTTCGCAATACAGTCTACACGCGCAGCGCAATTGGCGCGGACGAGACGCG 401
QY 111 sGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAl 131
Db 402 GCGCAGCAGCGCGCGCAATCAATGAATCGATGATGAAAACGTC----- 444
QY 131 aLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeuTyrSerSerGlyAs 151
Db 445 -----AAAGCTGTGAAATCAGCAAGGCTCAAACTCGGTTCGAACAAGCGCGCGC 497
QY 151 nValAlaGlyLeuVal-----AspValAlaAspGlyLysIleProGluLys-- 166
Db 498 ATTGGCGGCTCGGTGCGATTTTCAACCAAAACCCGCGACGATGTTATCGGGAAGGCG 557
QY 167 -----MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSe 182
Db 558 GCAGTGGGGCATTCAGATGATAAACCCGCTATTCCGCGAAA-----AACCG 602
QY 182 rGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheVa 202
Db 603 GGGGCTTACCAATCCATCCGCTGCGGGG---CGCATCGCGGTGCGGAGGCTTTGCT 659
QY 202 lleuHisThrGluGlyLeuTyrArgLysSerGlyAspTyr----- 215
Db 660 GATCCACACCGGG-----CGCGCGCGGGGAAATCCGCGCCACGAAAGATGCGAG 710
QY 216 -----AlaValProArgTyrArgAsnLeuLysArgLeuProAspSer----- 229
Db 711 ACGCGGCGTTTCAGAGCTTTTAAACAGGCTGTGCGGTGGAAGACAGCAGCAATACGCCTA 770
QY 229 ----- 229

```

Db 771 TTTTCATCGTTAAAGAAAGTGCMAAAACCGGAGTATTGAAACGTGTAAAGCGAATCCGAA 830
Qy 230 -----Pr 230
Db 831 AAAAGATGTTGTCGGCAAGACGAAACGTCACAAACGGTTTCCACCGGAGACTACACGGGTCC 890
Qy 230 oArgArgphe-----AlaasnGlyGlnHisArgAlaValLeu-----G1 243
Db 891 CAACCGCTTCTCGCGGATCCGCTTTTATACGAAAGCCGCTGCTGCTGTTCGCCCGCGG 950
Qy 243 YTPArg-----LysArgPheTyr-----ArgArgTh 252
Db 951 TTTTCGTTTGAATAAGCGCACTACATCGCGCGGCACTCGAACACACGCAACAAAC 1010
Qy 252 rTyrSerAspArg----- 256
Db 1011 TTTTCGACCGCGGATATGACGTTCCGGCATTCCTTGACCAAGCGGTTTTTGTGACAA 1070
Qy 257 -ArgAspGlnTyrGly---LeuPro-----AlaHisSerHisGluTyrAs 270
Db 1071 TAAATAAACAGCGCGGTTCTTTGCGCGGTAAACGGCAATACGCGGCAACCAAAATACGG 1130
Qy 270 pAspCysHisAlaAsp----- 1176
Db 1131 CGGACTGTTTACCAACCGCGAAACGTCGCTGCTGGCGCGGAATACGTTACGGCGCT 1190
Qy 276 eileTTPGlnLysSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuTh 296
Db 1191 GTTTTACGACGAGACGACACCAACCAAGCGCTACCGTTTGGAAATATCTATACCAATGC 1250
Qy 296 rGluGluAsp-----ValAspTyrAspAsnProGlnLysSerCy 309
Db 1251 CGATAAGACACTTGGCGGATATGCGCGCTCTCTTACGACCGCGGCGGCTGCGTTT 1310
Qy 309 sGlyPheHis-----AspAspAspAlaHisAlaHisAl 321
Db 1311 GGATAATCATTTTCAGCAGCGCACTGTTCTGCGCGCGTTCGCAACAAATATTCGCCGCC 1370
Qy 321 aHisAsnGlyLysProTrpLeuAspLeuArgAsnLysArgTyr----- 335
Db 1371 GAGTGCCGCAAGCGGTTTCTCTATTACAAATCGATCGCTGATTTTACGGGGAAAGCCA 1430
Qy 336 ----GluLeuArgAlaGluTrpLysGlnProPhePro----- 346
Db 1431 CAGGCTCTTCGAGCGGCATTCMAAAATCCTTCGATACGCCCAAAATCCGCCAACACCT 1490
Qy 347 -----GlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis-- 361
Db 1491 GAGCGTGAATCTCGGTTTGACCGCTTTGGCTCTAATCTCCGCCATCAGGATTATTATTA 1550
Qy 362 ----HisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsn-GlnThrGlnA 380
Db 1551 TCAACATGCCAACCGCGCTATTGCTGAACACACGCCCTCTCAAAACACACGCAAAAAAAT 1610
Qy 380 snAlaArgIleGluLeuArgHisGlnProIleGly-----ArgL 393
Db 1611 CAGCCCCAACGGCAGTGAACAGCCCTATTGGTCTACCATAGCAGGGGAAATGTCGT 1670
Qy 393 euLysGly-----SerTrpGly-----ValGlnTyrLeuGlyLysSers 407
Db 1671 TACGGGCAAAATCTCGCGCTTTGGGCAACAATATTATACGACTGACGCGCGCGCAGCAT 1730
Qy 407 erAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValG 427
Db 1731 CAACGGTAAAGCTATTACGGC-GCAGTTCGG-----GACAAAT----- 1767
Qy 427 lnHisTyrSerPhePheGlyValGlnAlaAsnTrpAspAsnPheThrLeuGluGlyG 447
Db 1768 -----GTCCTTTGGCGAGGTGGCGGATGTC----- 1794
Qy 447 lyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArgG 467
Db 1795 -----GGCGCGGCTTTCGCTACGAC----- 1815

Qy 467 luAsnTyrTyrLysGlnProLeuProAspLeuGlyAla---HisArgGlnThrAlaArgS 486
Db 1816 -----TACCGCAGCAGCATTCGAGCAGCGAGCGTTTCCACCGCAGCAGCCGCA 1867
Qy 486 erPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThr----- 504
Db 1868 CTTTGCTTCGGAACCGCGGCGCATCTCTCAAACTACCGACTGCTGCTGATTTGACTTACC 1927
Qy 505 --AlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisV 524
Db 1928 GCACCTCAACCGGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1987
Qy 524 alaAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnI 544
Db 1988 TTTCAAAGCAAGCGCGTCAAAATC-----GATCCGGAATAATCGTTCAACA 2032
Qy 544 leGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgA 564
Db 2033 AAGAAGCGCGCATCTGCTGTTTAAAGCGATTTTCGCACTTGGAGGCAAGTTGGTTCAACA 2092
Qy 564 snArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsnAspGly----- 578
Db 2093 ATGCTACCGCATTTGATTTGCTCGGGGTTATGAAGCGCAAAATTAAGAAGCGCAAGAA 2152
Qy 579 ----ArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnG 597
Db 2153 AAGCCAAAGCGACCGCGCTTCTCAATCCCAAGCGCGGATTTACCGGCATCAATA 2212
Qy 597 lnSerGly----- 599
Db 2213 TTTTGGCAAAATCGATTGGAAACCGCGTATGGGATAAATTCGCCGAAGTTGGTTACTTA 2272
Qy 600 -----AlaAspPheTyrGlyA 605
Db 2273 CATTTGCTATATATCGTGTCTGTCGCGACATCAAAAAACGCGCAGCCGCGATA 2332
Qy 605 laGluGlyGluIleTyrPheLysProThrPro---ArgTyrArgIleGlyValSerGlyA 624
Db 2333 TTCAATCAGATCTGCTGTGATGTCCTCAACCTCGCGCTATGCTGCTGCTGCTGCTGCTG 2392
Qy 624 spTyrValArgGly-----ArgLeuLysAsnL 633
Db 2393 ACCAACCGGAAGGCAAAATGGGTGTGAACGTATGCTGACTATTTCAAAGCCAGGAAA 2452
Qy 633 euProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspA 653
Db 2453 TCACAGAGTTG-----TTGGGCGAGCGCGGCTTTGCTCAACGCGCAACA 2494
Qy 653 spGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuT 673
Db 2495 GCCGCAATACAAAAGCCACCGCGCGGTACCCGCCCTTGGTATATT----- 2541
Qy 673 hrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaA 693
Db 2542 -----GTGACGTGTCC---GGTTATTACCGTT----- 2568
Qy 693 rgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgA 713
Db 2569 -----AAAAAACACTTTACCTTCCGTCGCGCGGTGATCAACCTCTCTCA 2611
Qy 713 snThrArgTyrGlyGluTyr---AsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnS 732
Db 2612 ACTACCGCTATGTTTACTTGGGAAATGTGCGGCAAACTGCGCGCGCGCGCGTCAACCAA 2670
Qy 732 erValTyrAlaHis-SerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThr 751
Db 2671 -----CACAAAATGTGCGCGTTTACACCGATATGCGCGCCCC---GGTTCGCAACTACACA 2724
Qy 752 GlyGlyValAsnValLysPhe 758
Db 2725 TTTAGCTTGGAAATGAAGTTT 2745


```
QY 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys----- 534
Db 12210 GATTTTAAATAAGTTCACAGTTACGATATAGTACAAAAGATCCAGTCACCGGTTA 12269
QY 535 -----HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu 548
Db 12270 AACACTTTTATTATGATATAGTGTAGGCAAAATATCAGAGGGGTGGAGCTTCAACT 12329
QY 549 GlyTyrGlu---GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
Db 12330 CAGATTCCTGTGTATGATAAATGGCATGTATCT-----GCA 12365
QY 568 AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAsp 587
Db 12366 AACTAT-----ACATTACTGACTCTCGT-----CGAAAAAGTATGACGAA 12407
QY 588 SerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGly 607
Db 12408 AGT-----CTCATGGCAAGTCGCTGAAGGGGAACCT--- 12440
QY 608 GluIleTyrPheLysProThrProArgTyrArgIleGly-----ValSerGlyAspTyr 625
Db 12441 -----CTGAAAGAACTCCACAGATCGACCAATGCAAAACTGGAATGGGATTAC 12491
QY 626 ValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn 645
Db 12492 ACTCAGGAT---ATTACATTTTATTCTCATCTCTG-----AATTATACGGGAAA 12536
QY 646 ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaIleArgLeuGly 665
Db 12537 CAAATCTGGCGAGCAAA-----AGAAATGGTGCTAAGTTCCTCCCGCTGTAATGGA 12590
QY 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db 12591 TTC-----ACATCTATGGATATTGGTCTAAATTACCAG----- 12623
QY 686 PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsn 705
Db 12624 ATTCTGCCAGACACGCTGATTAATTTTGGC-----GTCTTAAC 12662
QY 706 LeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsnTyrValLysAl 725
Db 12663 GTCAC-AGACAGAAAGACGAGATATCGATACATGTATGTTAGTTCGACGCTC----- 12716
QY 725 aAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProG 745
Db 12717 -----GA 12718
QY 745 nMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 12719 TGAAGACGCGCTTATGGGCTAATGTAAAGATATCCTTC 12758
```

RESULT 36

```
US-10-043-344-1
; Sequence 1, Application US/10043344
; Publication No. US20030088086A1
; GENERAL INFORMATION:
; APPLICANT: Loomis, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murdin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 1
; LENGTH: 4699
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(1940)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1957)..(4696)
US-10-043-344-1
Alignment Scores:
Pred. No.: 6,278-06 Length: 4699
Score: 155.00 Matches: 188
Percent Similarity: 32.33% Conservative: 144
Best Local Similarity: 17.41% Mismatches: 309
Query Match: 3.84% Indels: 344
DB: 14 Gaps: 48
```

US-09-936-377-2 (1-758) x US-10-043-344-1 (1-4699)

```
QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 1961 ACTAAAAAACCCCTATTTTCGCCCTAAGTATTATTCTTGTCTTTTAATTCATGCTATGTA 2020
QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2021 AAAGCAGAAACTCAAAGTATAAAAGATACAAAGAAAGCTATATCATCTGAAGTGGACACT 2080
QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 2081 CAAAGTACAGAGATTCAGAATTAGAACTAATCTCAGTCACTGCAGAAAAGTTAGAGAT 2140
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2141 CGTAAAGATAAATGAAGTA-----ACTGGACTTGGCAAAATATATAAAACTAGT 2188
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2189 GAAAGATATCAGCCGAGAACAAAGTATTAAATATTCGTGATCAACAGCTATGATCCAGG 2248
QY 83 IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
Db 2249 ATTTCAGTGTAGAACAGGTCGCGTGCAGAGTTCTGGATATTCTATTCTGTGTATGGAC 2308
QY 102 GlyArgArgIleLysValLeu----- 108
Db 2309 AGAAATAGAGTTGCTTTTATTAGTAGTGTTTACCTCAAACGCAATCTTATGTAGTGCAG 2368
QY 109 -----AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 2369 AGCCCTTTTATGCTCGTTCAGGATATTCGGCAGATATTAATTAATTAATTAATTAAT 2428
QY 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro 142
Db 2429 GAAAAATGTA-----AAGCCCTCGAAATAAGCAAGGGGGG 2464
QY 143 ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal----- 156
Db 2465 AGTTCCTCTGAGTATGTAATGAGCACTAGTGGTCTGTGTAAACATTTCAAAGCAAAATCA 2524
QY 157 -----AspValAlaAspGly-----LysIleProGluLys 166
Db 2525 GCAGCCGATATCTTAGAAGGAGCAAAATCATCGGGAATTCAAACTAAATAATGCTTATTC 2584
QY 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186
Db 2585 AGCAAAATAAAGCCTTTTACCCTATCTTACCTGTAGTGTAGCAGAAAACAAGGTGGATTGAA 2644
QY 187 LysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn----- 200
Db 2645 -----GGGTGCGCATTTTACACTCACCGAATTCATTTGAATGAACCAAGTC 2699
```

QY 201 -----PheValLeuHisThrGlu--- 206
Db 2690 CATAAGATGCATTAAGCGGTGCAAGTATGATCGATTATCGCCCAACAGAGGAT 2749
QY 207 ---GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225
Db 2750 CAATCGCATACCTTGTGATGCAAGATGAGTGTGTAGATGGTTATGACAAGTGCTAAACT 2809
QY 226 LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu--- 242
Db 2810 TCACCCAAACGACCTCGACTTATCCACCAAGAGAAACCGTAAGCGTTTCAGATTAT 2869
QY 243 -----GlyTyrPArgLys 246
Db 2870 ACGGGGCTAACCGTATCAAACTTAATCAATGAATATGAAAGCCAGCTCTGG--- 2923
QY 247 ArgPheTyrArgArgThrTyrSerAspArgAspGlnTyr-----GlyLeuProAla 264
Db 2924 ---TTTAAAGAGAGGATTATCATTTTCTGCAACACACTATATGTGTGTAT--- 2974
QY 265 HisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsn 284
Db 2975 -----TTTGAATTCACACACAAATAATTTGATATC----- 3004
QY 285 LysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsn 304
Db 3005 ---CGTGATATGACATTTCCCGCTTATTTAAGGCCAACAGAGAGATTTACAA--- 3058
QY 305 ProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGly 324
Db 3059 -----AGTCGCCCTTTTATCCAAAGCAAGATTTATGTGATATCAACATTTGGT 3109
QY 325 LysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnPro 344
Db 3110 -----GATGGCAGAGCGGCTTAAATAT----- 3130
QY 345 PheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlu 364
Db 3131 -----GCAAGTGGGCTTTATTTTCGATGAACACCATAGAAAAACAG 3169
QY 365 LysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaAargIleGlu 384
Db 3170 CGTGTAGT-----ATTGAATATATTTACGAAATAATAAGAACAAAGCGGCATCATTCAC 3223
QY 385 LeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGln 404
Db 3223 ----- 3223
QY 405 LysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn 424
Db 3224 ---AAAGCGGTGTTAAGTCTTAATCAACAAACATCATCTTGACAGTTATATGCGACATA 3280
QY 425 LysVal-----GlnHis 428
Db 3281 CGCATTCAGCTTTTATCCATATCCAAAGTAAGATTGCGGCCCAACACTGTGATAACCT 3340
QY 429 TyrSerPhePheGlyValGlu----- 435
Db 3341 TATTCATACATCATCTGTAGATAAATGTTTATAAAGAAAAACATAACATGTTGCAATTG 3400
QY 436 -----GlnAlaAsnTrp----- 439
Db 3401 AATTAGAGAAAAAATTCACAAATTTGGCTTACTCATCAAAATGGCTTCAATCTTGGT 3460
QY 440 ---AspAsnPheThr-----LeuGlu 445
Db 3461 TTTGATGACTTTACTTCCGCACTTCAGCATAAAGATTATTTAACTCGACGTGTATCGCT 3520
QY 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAsp 465
Db 3521 ACGGCAAGTAGTATTTTCAGAGAAACCGTGTGGAAGCAAGAAATGGTTTA----- 3571

RESULT 37

QY 466 ArgGluAsnTyrTyrLysGlnProLeuPro-----AspLeuGly 478
Db 3572 CAATCAAGTCCTTACTTATATACCAACACCAAAAGCAGAGTTGTTAGCAGAGATCTTTGT 3631
QY 479 AlaHisArgGlnThrAlaAargSerPheAla-----LeuSerGly--- 491
Db 3632 AATTATCAAGTAAGTCCTTAATACAGTGACTGTAAGTCGGTAAATTAAGGGA 3691
QY 492 AsnTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 3692 AATTATTTTCGACGACGCAATAATATGCGCATTA----- 3727
QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 3728 -----GGGAATACGTT-----GATTTAGGTTTA 3751
QY 532 GlyAsnLysHis---LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly--- 549
Db 3752 GGTATGAGGTATGACGTATCTCTGACAAAGCTAATGAATCAACTATTAGTTGGTAAA 3811
QY 550 TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
Db 3812 TTTAAAAATTTCTTTGGAAATCTGTTATTTGCTATAAAACCAACGGAATGCTTGATCTT 3871
QY 570 IleTyrAlaGlnThrLeuAsnAspGly---ArgGlyProLysSerIleGluAspAspSer 588
Db 3872 TCTTAT-----CGCCTTTCTCTGGAATTTAGAAATCTTAGTTT-----GCT 3913
QY 589 GluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu 608
Db 3914 GAAATGATGTTGGTGGGTATGGTGGCAAGGATACCGATGTTTATATAGGTAAA----- 3967
QY 609 IleTyrPheLysPro-----ThrProArgTyrArgIleGlyValSerGlyAsp 624
Db 3968 -----TTTAAGCCTGAAACATCTCGTAACCAAGATTTGGTCTCGCTCTAAAAGGGAT 4021
QY 625 TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly 644
Db 4022 TTTGGTATATGATGATCAGTCATTT-----AGTAATGCTTATCGA 4063
QY 645 AsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAargLeu 664
Db 4064 AATCTTATCGCCTTTGCTGAAGAACTTAGTAAAAATGGAACCTACTTGGAAAGGCAATTAT 4123
QY 665 GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAsp 677
Db 4124 GGATATCATATATGCACAAATGCAAAATGATTTGGCGTAATATATACTGCGCAATTAGAT 4183
QY 678 AlaAsn-----LeuAspTyrTyrArgValPheAlaGlnAsnLys 690
Db 4184 TTTTATGTTTATGGAACGTTATCCCTACGTTGTTGATGCAACATTTGCTTATAACCGA 4243
QY 691 LeuAla-----ArgTyr 694
Db 4244 GTAAAGTTAAAGATCAAAAAATCAATGCTGTTTGTCTCGTAAGCAGTTTATTTATTT 4303
QY 695 GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThr 714
Db 4304 GATGCCATTCACGCCAGCGGTATATCATGTTTGTAGGTATGATCATCCCAAGTAATACT 4363
QY 715 ArgTyrGlyGluTyrAsnTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyr 734
Db 4364 -----TGGGAATTAAG-----ACAATGTTT 4384
QY 735 AlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrGly----- 752
Db 4385 ACTCAATCAAAAGCAAAATCTCAAAATGAATTTGCTAGGAAAACGTCGATTTGGGTAAACAT 4444
QY 753 GlyValAsnValLys 757
Db 4445 TCAAGGAATGTAATA 4459

```

US-10-043-344-105
; Sequence 105, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murdin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 105
; LENGTH: 5144
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
US-10-043-344-105

Alignment Scores:
Pred. No.: 9.25e-06 Length: 5144
Score: 154.00 Matches: 196
Percent Similarity: 31.72% Conservative: 125
Best Local Similarity: 19.37% Mismatches: 331
Query Match: 3.82% Indels: 360
DB: 14 Gaps: 50

US-09-936-377-2 (1-758) x US-10-043-344-105 (1-5144)
Qy 5 ThrLeuIysProIle-----ValLeuSerIleLeuIleAseThrProLeu 20
Db 2138 ACTAAAAACCCCTATTTTCGCCTAAGTATTATTTCCTGCTTTTAATTCATGCTATGTA 2197
Qy 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2198 AAGCAGAAACTCAAGATATAAAGATACAAAAGAAGCTATATCATCTGAAGTGGACACT 2257
Qy 33 -----LeuGluThrValThrValValGly----LysSerArgPro 44
Db 2258 CAAAGTACAGAAGATTCAGAAATTAGAACATATCTCAGCTCACTGCAGAAAAAATAAGAGAT 2317
Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2318 CGTAAAGATATGAAGTA-----ACTGGACTTGCCTGCAAAATATAAAGACGAGT 2365
Qy 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2366 GAAAGTATCAGCCGAGAACAAAGTATTAATATTCGTGATCTTAAACACGCTATGATCCAGGC 2425
Qy 83 IleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal----IleArgGlyGlnThr 101
Db 2426 ATTTCAAGTTGTAGAACAAAGGTCGGGTCGCAAGTTCTGGATATTCCTATTCGTGGTATGGAC 2485
Qy 102 GlyArgArgIleLysValLeuAsnHisIleGlyGluThrGlyAspMetAlaAspPheSer 121
Db 2486 AGAAATAGAGTT----- 2497
Qy 122 ProAspHisAlaIleMetValAspThrAlaIleuSerGlnGlnValGluIleLeuArgGly 141
Db 2498 -----GCTTTATTAGTAGATGGTTTACTCAACGCAACTCTTATGATAGTCAAGC 2548
Qy 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 2549 CTTTATAGTTGCTCGTTCCAGGATATTCGGCACTGGTGCATTAATGAATGAATATGA 2608
Qy 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179

```

QY 440 ----- AspAsnPhe ----- 442
Db 3617 GTCTCAATCTGGTTTGGTACCTTTACTTCAGCGCTTCAGCATAAAGATTATTTAACT 3676
QY 443 ----- ThrLeuGluGlyValArgVal 450
Db 3677 CGACGTGTTACCGCTACGCAAAATATTATTTTCAGGACAGTTGCTGTTAAAGCAAAAT 3736
QY 451 GluLysGlnLysAlaSerIleArgTyrAspLys ----- 461
Db 3737 GGTTCAGAAAACAACCTTACTTATCTACCAAAACCAAAAGTAGATTTTGTAGGACAAGAT 3796
QY 462 ----- AlaLeuIleAsp 465
Db 3797 CATTCGTAATTATAAAGTAGTCTCTTAATTACAGCGACTGTAAAGTGGCGTTAATTAAA 3856
QY 466 ArgGluAsnTyrTyr ----- LysGlnProLeuProAspLeu 477
Db 3857 GGGAAAAATTTATTTATTCGACGCAACCAATATATGCGAATAGGGAATACATTTGATTTA 3916
QY 478 Gly ----- AlaHisArgGlnThrAla ----- 484
Db 3917 GGTTCAGTATTCGCTATGACGTATCTCGTACAAAAGCTAATGAATCACTATTAGTGTT 3976
QY 485 ----- ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu 501
Db 3977 GGTAAATTTAAAAATTTCTCTGGAATACTGCTATTTGTCATAAAACCAACGGAATGGCTT 4036
QY 502 SerLeuThr ----- AlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla 519
Db 4037 GATCTTTCTATCGCGCTTTCTACTGGAATTTAGAAATCCCTAGTTTTCGCTGAAATGTAT --- 4093
QY 520 HisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGlu 539
Db 4094 --- GGTTCGCGTATGCTGGCAATAATAGCGATGTTTATGTAGGTAAATTTAAGCCTGAA 4150
QY 540 ArgSerAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeu 559
Db 4151 ACATCTCGTAACCAAGAGTTTGCTCTCGCTCTAAAGGGGATTTTGCTAATATTTGAGATC 4210
QY 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIle --- TyrAlaGlnThrLeuAsn --- 576
Db 4211 AGTCATTTTAGTAATGCTTATCGAAATCTTATCGCCCTTTGCTGAAGAACTTAGTAAAT 4270
QY 577 --- AspGlyArgGlyProLysSerIleGluAspSerGluMetLysLeuValArg 594
Db 4271 GGAACCTACTGGAAGGCAATTTATGATATCATATATGACAAAATGCAAAATTTAGTTGC 4330
QY 595 TyrAsn --- GlnSerGlyAlaAspPheTyrGlyAlaGluGlyLeuTyrPhe --- 611
Db 4331 GTAAATATACTCGCAATTAGATTTTAAATGTTTATGTTATGGAACGTTATCCCTACGTTGG 4390
QY 611 --- 611
Db 4391 TATGCAACATTTGCTTATACCGAGTAAAGTTAAAGATCAAAATCAATGCTGTTTG 4450
QY 612 --- LysProThrProArgTyrArgIleGly 620
Db 4451 GCCTCCGTAAGCAGTTATTTATTTGATGCGCATTCAGCCAGC --- GCTTATATCAITGGT 4507
QY 621 ValSerGlyAsp --- TyrValArgGlyArg 629
Db 4508 TTAGGCTATGATCATCCAGTAATATCTGCGGAATTAATACATGTTTACTCAATCAAAA 4567
QY 630 LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db 4568 GCAAAATCTCAAATGATTTGCTAGGACACGT --- GCATTCGGGTAC --- 4612
QY 650 AlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLys 669
Db 4613 --- AATTCAGGAATGTAAATCAACAAGAAAACCTTACTCGGGCATGGCATATC --- 4663

QY 670 AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsn 689
Db 4664 --- TAGATGTATCGGGTTATTCATATTCGCGAAT 4693
QY 690 LysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsn 709
Db 4694 AAA --- AATATTATGCTTCGATTAGGATATAT 4723
QY 710 TyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLys --- AlaAspAsnLeu 728
Db 4724 AATTTATTCAACTATCGCTATGTTACTTGGGAAGCGGTGCGTCAAAACAGACACAGGTGCG 4783
QY 729 LeuAsn --- GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMet 746
Db 4784 GTCATCAACATCAAAATGTTGGTAGCTATATCTCCTACGCGAGCATCA --- 4831
QY 747 GlyArgSerPheThrGlyValAsnValLysPhe 758
Db 4832 GGACGAACTATACCTTAACATTAGAAATTC 4867
RESULT 38
US-09-332-226-1
; Sequence 1, Application US/09332226
; Patent No. US20020025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,226
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/363,124
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3286 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:


```

Db      2185  ---AGTGGCGGATGTCGGCGGGCTTCGCTACGAC----- 2220
Qy      468  AsnTyrTyrLysGlnProLeuProAspLeuGlyAla---HisArgGlnThrAlaArgSer 486
Db      2221  -----TACCGCAGCAGCATTCGACGACGCGCGTTCACCGCAGCGCAGCGCACC 2274
Qy      487  PheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeuThr----- 504
Db      2275  CTGCTCTGGAACCGCGCATGCTCCTCAAACTCCGACATGGCTGGATTTGACTTACCGC 2334
Qy      505  AlaSerHisGlnGluArgLeuProSerThrGlnGlnLeuTyrAlaHisGlyLysHisVal 524
Db      2335  ACTTCAACCGGCTTCCGCTCCCTCGTTTCGGAATGACGCG----- 2379
Qy      525  AlaThrAsnThrPheGluValGlyAsnLys-----HisLeuAsnLysGluArg 540
Db      2380  -----TGGCGTGGCGGATATAATAAAGCCGTCAAATCGATCCGGAATAA 2427
Qy      541  SerAsnAsnLeuGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAla 560
Db      2428  TCGTTCAACAAGACCGCGCATGCTGTTTAAAGCGGATTTTCGCAACTTCGGAGCGCAAGT 2487
Qy      561  LeuTyrArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsnAsp 577
Db      2488  TGGTTCAACATGCTACCGCGATTTGATTGTCGGGGTATGAAGCGCAATTAAGAC 2547
Qy      578  GlyArg-----GlyProLysSerIleGluAspSerGluMet----- 590
Db      2548  GGCAAGAACAAAGTCAAAAGCAACCGCGCTTACCTCAATGCCAACGCGCGGATTACC 2607
Qy      591  -----LysLeu----- 592
Db      2608  GGCATCAATATTTGGGCAAAATCGATTGGAACGCGGTATGGGATAAATGCCCGAAGGT 2667
Qy      593  -----ValArgTyrAsnGlnSerGlyAlaAsp 601
Db      2668  TGGTATTCACATTTGCTATATCGTGTCCGTGTCGCGCATCAAAACGCGCAGAC 2727
Qy      602  PheTyrGlyAlaGluGlyGluIleTyr-----PheLysProThr----- 614
Db      2728  CGCACCGATATTCATACACATGTTGATGCCATCCAAACCTCGCGCTATGTCGTCGCG 2787
Qy      615  -----ProArgTyrArgIleGlyValSerGlyAspTyrTyrValArgGlyArg 629
Db      2788  TCGGCTATGACCAACCGAAGCAATGCGCGTGAACGGTATGCTGACTTATTCCTCAA 2847
Qy      630  LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db      2848  GCCAAGGAATCACAGAGTTG-----TTGGCGACGCGGCTTTGCTC 2889
Qy      650  AlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLys 669
Db      2890  AACGGCAACAGCGCAATACAAAGCCACCGCGCGTACCCGCTTGGTATATT--- 2946
Qy      670  AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsn 689
Db      2947  -----GTGACGTGTCC---GGTATTACCGTT----- 2973
Qy      690  LysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsn 709
Db      2974  -----AAAAACACTTCACCTCCGTCGCGCGGTGATC 3006
Qy      710  TyrArgArgAsnThrArgTyrGlyGluTrp---AsnTyrTyrValLysAlaAspAsnLeu 728
Db      3007  AACCTCTCAACACCGCTATGTTACTTGGGAAATGTGGGCAAACTGCCGCGCGCGCA 3066
Qy      729  LeuAsnGlnSerValTyrAlaHis-SerSerPheLeuSerAspThrProGlnMetGlyArg 748
Db      3067  GTCAACCAA-----CACAAAATCTGGCGTTTACACCGGATATGCGCGCCCCC-GGCG 3119
Qy      748  gSerPheThrGlyGlyValAsnValLysPhe 758

```

```

Db      3120  CAACTACACATTTAGCTTGGAAATGAAGTTC 3150
RESULT 39
US-09-332-226-3
; Sequence 3, Application US/09332226
; Patent No. US20020025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: InClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,226
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/363,124
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3537 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria meningitidis
; STRAIN: FAM18, FAM20, B16B6, group X and group W135
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 721..3450
; FEATURE:
; NAME/KEY: mat-peptide
; LOCATION: 793..3447
US-09-332-226-3
Alignment Scores:
Pred. No.: 2,4e-05
Score: 148.00
Percent Similarity: 32.20%
Best Local Similarity: 20.46%
Query Match: 3.67%
DB: 9
Length: 3537
Matches: 204
Conservative: 117
Mismatches: 324
Indels: 355
Gaps: 48

```

```
US-09-936-377-2 (1-758) x US-09-332-226-3 (1-3537)
Qy 14 LeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeu 33
Db 766 TTAATGACCGCGTTCGCCGTTTATGACAAATGTGCAAGCCGCAACAGCAGCAGAAAA 825
Qy 34 GluThrValThrValGlySerArgProArgAlaThrSerGlyLeuLeuHisThr 53
Db 826 CAGTTGGATACCATACAGGTAAAGCCAAAGAAACAGAAAAACCCCGCGATACGAAGTA 885
Qy 54 SerThrAlaSerAspLysIleSerGlyAspThrLeu---ArgGlnLysAlaValAsn 72
Db 886 ACCGGCTGGCAAGTTGGTCAAGTCTCCGATAGCTAAGTAAAGAACACAGGTTTGAAT 945
Qy 73 LeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGlyVala 92
Db 946 ATCCGAGACCTGACCCGTTATGATCCGGTATTGCCGTGGTGCAGAACAGGGTCGGGGCGCA 1005
Qy 93 SerAlaProVal---IleArgGlyGlnThrGlyArgArgIleLysVal----- 107
Db 1006 AGTTCGGCTATTCAATACGGCGATGGATAAACCCGCTTCCTTAACGGTAGACGGC 1065
Qy 108 -----LeuAsnHisGly 112
Db 1066 GTTTCGCAATACAGTCTACACCGCGCAGCGCATTTGGGTGGGACGAGACGGCGGT 1125
Qy 113 GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
Db 1126 AGCAGCGCGCAATCAATGAAATCGATGATGATAAAACCCGCTTCCTTAACGGTAGACGGC 1164
Qy 133 SerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuLysSerGlyAsnVal 152
Db 1165 ---AAGCCGTTGAAATCAGCAGGGTTCGAATTCATCAGATACGGAACGGCGCATG 1221
Qy 153 AlaGlyLeuValAsp-----ValAlaAspGlyLys--- 162
Db 1222 GCAGTTCGGTCGATTTCAAAACCAACACCGCAGCCGCAATTTATCGGAGAGGGAACAG 1281
Qy 163 ---IleProGluLysMetProGluAsnGly-----ValSerGlyGluLeu 176
Db 1282 TGGGCAATCAGAGTAAACCTGCTATTTCGGGAAAGACCATGCCCCGACGCAATCCCTT 1341
Qy 177 GlyLeuArgLeuSerSerGlyAsnLeuGluLysLeu----- 188
Db 1342 CGCTTCGGCAGCAGCAGCGCGCGCGGAGAGCCCTCTTATTATTATTAACGCGCGGT 1401
Qy 188 ----- 188
Db 1402 CGGGAATCCATGCGCATAAAGATCCGCGCAAGGCTGTGCAAGCTTCAACCGGCTGTG 1461
Qy 189 -----ThrSerGlyGly-----IleAsnIleGlyLeu----- 197
Db 1462 TTGACGAGGACAAAGAGGTGGCAGTCAGTCAGATATTTCATTGTGGAAGAAGAA 1521
Qy 197 ----- 197
Db 1522 TGCCCAATGGATATGCGCCCTGTAAACCAAGCTGAAAGAGATGCTCGTCAAGAT 1581
Qy 198 -----GlyLysAsnPheValLeu----- 203
Db 1582 GAGCGCAAAACCGTCAGCAGCAGGATTATACCGGCTCCAAACCGCTTACTTTCGGAACCG 1641
Qy 204 -----HisThrGluGlyLeuTyrArg----- 210
Db 1642 CTTGAGTATGGCAGCAATCATGGCTGTTCGACCGGGTTGGCATTTGGACAAACCGCAT 1701
Qy 211 -----LysSerGlyAspTyrAlaVal 217
Db 1702 TATGTCGGAGCGTTCGAAAGTACGACGACACCTTTGTATACACGGGATATGACTGTT 1761
Qy 218 ProArgTyr-----ArgAsnLeuLysArgLeu----- 226
Db 1762 CCTGCTATTATTACCAGTGAAGATTATGTACCGGTTGCTGAAAGGCTTTGGCAATAT 1821
Qy 227 -----ProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db 1822 TCGGCGCATATAAGCAGAAAGCTGTTTTCAGGAGAGGG-CAGTACATTCACGG 1880
Qy 244 TrpArgLysArgPheTyrArgArgThr-TyrSerAspArg-----ArgAspGlnTyrG 261
Db 1881 TATCGGTACGG---TACCGCGTGTATTATGATGACGCCATATAAAACACCGTACGG 1937
Qy 261 YLeuProAlaHisSerHisGluTyrAsp---AspCysHisAlaAspIleIleTrpGlnL 280
Db 1938 GGTCAATATGTTTACCATAATGCTGATAGGATACCTGGGCCGAT----- 1983
Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
Db 1984 -----TACGCCCGCACTT----- 1995
Qy 300 LAspTyrAspAsnProGlyLeuSer-----CysGlyP 311
Db 1996 -TCTTATGACCGCAAGTATAGATTTCGACACCGTTTCGACGACGACGATTCGCT-- 2052
Qy 311 eHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuAr 331
Db 2053 -CAGCAGCGTTCGGATAAAATTCGCTCCCGACGGCAATAAACCGTATCTTCTATAA 2111
Qy 331 gAsnLysArg-----TyrGlu-----LeuArgAlaGluTrpLysGlnPr 344
Db 2112 ATCCGACCGGATGATTATGAAAGAAAGCCGAAACCTGTTCACGAGCATTTTAAAGGC 2171
Qy 344 oPhePro-----GlyPheGluAlaLeuAr 352
Db 2172 ATTTGATACGGCCAAATCCGTCACAAATTCGATATCAATCTAGGTACGACGCTTTAA 2231
Qy 352 gValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAs 372
Db 2232 CTCGCAATTGTCACACGAGTATTAT-----CTTCAAAACGCGATTCAGGC 2279
Qy 372 nPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg-----HisGlnProI 390
Db 2280 ATATGA-TTTGATAAACCCGAAAGCTTCGTTTCCCAACGGAAGCAACACCCGT 2338
Qy 390 e---GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuAspAsnLysValGlnHisT 409
Db 2339 ATAGGTGCTCTACGGCAAGCAGCGGT---CAATACATC----- 2375
Qy 409 uSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisT 429
Db 2376 -----GCCGATACCTGTTTCGCAATAACACCTATACAGACTGCACACCGGAATA 2428
Qy 429 r-SerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyVala 449
Db 2429 TCGGCGGCAACGTTATTATGACCGCTTCAGACAAATGTCGCTTTGGGC-----A 2479
Qy 449 rgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnT 469
Db 2480 GGTGGCGGATCTCGGACGACGATACGTACGATTAC-----CCGACGACCG 2527
Qy 469 yrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAla 489
Db 2528 ATTCGGAAGATAAGAGTGTCTCTACCGGCACTCACCGCAAC-----C 2569
Qy 489 euSerGlyAsn-----TrpTyrPheThrProGlnHisL 500
Db 2570 TTTCTTGAAGCGCGCGGTAGTCTCAAAACCTTTCACCTGGATG-----G 2614
Qy 500 ysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaH 520
Db 2615 ATTTGACTTATCGCGCTTCTACGGCTTCGCTCGCGCTGCTTCCGGAATGATGCT 2674
Qy 520 isGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluA 540
Db 2675 GGAGAGCGGGGAGTCTTTGAAAAACGTTGGAT-----CTGAAACCGGAAA 2719
```



```
Qy 540 rSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuA 560
Db 2720 AATCCTTTAATAGAGAGCAGGTATGTATTTAAAGGGGACTTCGGCAATTTGGAAGCA 2779
Qy 560 laLeuTyrArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsnA 577
Db 2780 GCTATTTCAACAATCCCTTCGCGACCTGATTCGATTCGGTTATGAAACCGCACTCAA 2839
Qy 577 spGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnG 597
Db 2840 ACGGCA-----2847
Qy 597 InSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgT 617
Db 2848 -----ACTTCGGCTTCGGCGAC-----CCCGAT 2872
Qy 617 yArgIleGly-----ValSerGlyAspTyrValArgGlyArgLeuLys- 631
Db 2873 ACCGAAATGGCCCAAAATGCACGGTAGTAGCCGGTATCATATTTGGTAAATCGATT 2932
Qy 632 -----AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnA 646
Db 2933 GGCACGGCGTATGGGCGGGTTCGCGACGGTGTATTCCACGCTTGCTAT---AACC 2989
Qy 646 rgProPheIleAlaGlnAspAspGlnAsnAlaProArg-----658
Db 2990 GTATCAAGGTCAAAAGTCGCGATATACGCCCGCAGCAGGCTTTGTAACTTCATATCTCT 3049
Qy 659 -----ValProAlaAlaArg-----LeuGlyPhe-----666
Db 3050 TTGATCCCGTCCACCTTCACGATATGTATTGGGTTGGGTTACGACCATCTCGACGGA 3109
Qy 667 -----HisLeuLysAlaSerLeuThrArgArg-----675
Db 3110 TATGGGCGCATCAATACGATGTTTACTTATTCAGGCAAAATCTGTTGCAAGACTGCTCG 3169
Qy 676 -----IleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysL 691
Db 3170 GCAGCAGCGCGTGTGTGAACGGTATGCCAATGCTAAAGAAAGCAGCATCACCGGAGCG 3229
Qy 691 euAlaArgTyrGluThrArgThrProGly-----HisHisMetLeuA 705
Db 3230 GGCCTGGTGTATGTCGGATGTTCCGGATATTACAATATCAAGAAACACCTGACCCCTGC 3289
Qy 705 snLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrp---AsnTrpTyrValL 724
Db 3290 GGCAGGTGTGTACAACTCTCAACTACCGCTATGTTACTTTGGGAAATATGCGGCAAA 3349
Qy 724 yAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHis-SerSerPheLeuSerAspThr 743
Db 3350 CTGCGCGCGCGCAGTCAACCAA-----CACAAAATGTCCGCTTACACCGGATATG 3403
Qy 744 ProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 3404 CCGCCCCC-GGCGGAACTACACATTTAGCTTGGAAATGAAGTTT 3447
```

RESULT 40

```
US-09-815-242-6035
; Sequence 6035, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6035
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2190)
US-09-815-242-6035
```

```
Alignment Scores:
Pred. No.: 6,03e-05 Length: 2190
Score: 141.50 Matches: 151
Percent Similarity: 32.36% Conservative: 93
Best Local Similarity: 20.03% Mismatches: 275
Query Match: 3.51% Indels: 235
DB: Gaps: 35
```

US-09-936-377-2 (1-758) x US-09-815-242-6035 (1-2190)

```
Qy 3 GlnThrThrLeuLysProIleValLeuSer-----IleLeuLeuIleAsnThrProLeu 20
Db 37 CAAGCCATCACCAACCGTCACCTACTTCCGGTTGCATAGCAGCTGCACATTACCTTCT 96
Qy 21 LeuAla---GlnAlaHisGluThrGluGlnSerVal-----GlyLeuGluThrVal 36
Db 97 GCGGCTTTTGTGTCACCCACCTGAGAAACGGTGATTGTTGAGGGTTTCAGCCAGCT 156
Qy 37 ThrValValGlyLysSerArgProArgAlaThrSer-----GlyLeuLeuHisThr 53
Db 157 CCAGATGATGGCGAAATGATTACAGCGTAACGCTCTACCTCTCGCGGTACCAAAATGCAG 216
Qy 54 SerThrAlaSerAsp-----LysIleIleSerGlyAspThrLeuArg---67
Db 217 ATGACTCAACGTGATATTCCTCAGTCGGTCACTATTGTTAGCCAGCAGCGGATGGAAGAT 276
Qy 68 GlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGln 87
Db 277 CAGCAGTTACAAACGCTGGCGAAGTATGAGAAACACACCTGGGATCAGCAAAATTCAG 336
Qy 88 TyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysVal 107
Db 337 GCGGATTCGATCGTGTCTCTTTATTATTCGCCGGA-----TTCCAGATC 381
Qy 108 LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis-----124
Db 382 GATAACTAT-----ATGGTTGATGGTATCCCACTATTGTTGATTCG 423
Qy 125 -----AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu 139
Db 424 CGCTGGAATCTGGCGACGCACTTTCTGATATGGCACTCTTTCAACCGCTAGAGTAGNG 483
Qy 140 ArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 484 CGTGGCGCAGCAGGACTCATGCCGCGGCGGTAACTCCATCTCGCGCAATTAATATGGTT 543
Qy 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179
```

1435 -----CGGTTGATACGCTGACTTACAGCATGGAGAAAAAACCCACACCGCCTTAC 1495
475 ProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyr 494
1486 GCTGGTCTG-----GTGTTTGACATCAATGACAACTGG--- 1518
495 PheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThr 514
1519 -----TCGACCTACGCCAGCTATACCTCTATTTTCCAGCCCA 1557
515 GlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys 534
1558 AATGATCGTCACAGTTTCAGGCAATATCTGGCTCCAATCACC----- 1599
535 HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu-----GlyTyrGluGly 552
1600 -----GGTAACACTACGAGCTGGGTCTGAAATCGGACTGGATGAAT 1641
553 AspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAla 572
1642 AGCGTCTGACCAACCAGTTAGCCATCTTCGGT----- 1674
573 GlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeu 592
1675 -----ATTGAGCAGGATAATGTCGCTCAG--- 1698
593 ValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLys 612
1699 -----TCCACCGGTACACCTATCCCGCAGCAACGGCGAAACCGCTATAAA 1746
613 ProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsn 632
1747 -----GCGTGTGATGGGACAGTCAGTAAAGGGGTGGAAATTGAA 1785
633 LeuProSer-----LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAla 650
1786 CTCACGGCGCAATTACCGCAACTGGCAGCTGACATTTGGCGGCAACGGCTATATTGCA 1845
651 GlnAspAspGln-----AsnAlaProArg 658
1846 GAGGATAACGAAGAAACGCCGTTAATCTCTATCTGCCACGC 1887

Search completed: November 17, 2003, 21:24:16
Job time : 2666 secs

544 CGAAACACGCGCAGCAGTGGTGAATTTAAAGCGCATGCTCGCGCGAAATACGGT----- 597
180 LeuSerSerGlyAsnLeuGluLys-----LeuThrSerGly 191
598 -----AGCTGGAAACAAAGAACCGTATGTCGGCGATTACAAAGCCCACTCACGGAAGAC 651
192 GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys 211
652 GGT-----AAAATCCGCGCGCAATTGTGCGCGGTAC--CAG 687
212 SerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArg 231
688 AATAACGACTCATGGCTGACCGCTAC-----AACAGTGAAGAAG 726
232 ArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPhe----- 248
727 ACCTTCTTCGCGCATTTGCGATGCTGATTAATAGCCCTACCTGGCGCGGTTTACCGCGCTGAAT 786
249 -----TyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro----- 263
787 TACGAATATCAGCGCATTTGATGTTAATAGCCCTACCTGGCGCGGTTTACCGCGCTGAAT 846
264 -----AlaHisSerHisGluTyrAspAspCysHisAla-----AspIleIleTrp 278
847 ACTGATGCGCAGCAGCAACAGTTACGATCGCGCAGCAGTACCGCACCTGATGGCGGTAC 906
279 GlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu----- 294
907 AACGTAAGAAGATCAACAGGTCTTTATGACCTGAAAGCAGCAGTGTGCTGATACCTGG 966
295 -----LeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys 309
967 CAAGCGCACATGATGCCACCCACTCTGAAGTCGAATTTGACAGCAAAATGATGTATGTC 1026
310 GlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrpIleAsp 329
1027 -----GATCCCTATGTAACAAAGCGGATGGTATG----- 1056
330 LeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGlu 349
1057 -----CTGGTTGGCCATACAGTAATTTATGACCTGGCTGGTTGAT 1095
350 AlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAla 369
1096 -----TATGTCGGCGCACCGGTTGGAAACAGTGGCAAAAGTGAAGTTGATGCG 1143
370 ValGluAsnPhePhe----- 374
1144 CTGGATTGTTGCTGCTGACGGTAGTATGAATGTTGTTGCTGCTGACCAATCTAATGTTT 1203
375 -----AsnAsnGlnThrGlnAsnAlaAlaArgIleGluLeuArgHis 387
1204 GGTGGCAGTTACAGCAAAACAAACAAATCGTTACTTCAGTTCTATGGGCAACATCTTCCCG 1263
388 GlnProIleGlyArgLeu-----LysGlySerTrpGlyVal 399
1264 GATGAAATTTGGCAGTTTCTACAACTTAAATGGCAATTTCCCAAAACCGCTGGTACCA 1323
400 GlnTyrLeuGlyGln-----LysSerSerAlaLeuSerAlaThrSerGlu 414
1324 CAGAGCTGGCGGAGGAGGATACCAACATATGAAATCGTTATATGCTGCCACTCGTGTC 1383
415 AlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyVal 434
1384 ACCCTTCGCGATCGCTGCATCTG-----ATCCTCGGCGCA 1419
435 GluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
1420 CGTTATACCAACTGG----- 1434
455 AlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeu 474

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 17:07:47 ; Search time 2659 Seconds

(without alignments)

6928.467 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPVLISILLINTEL.....FLSDTPQMGSRFTGWNVKF 758

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cpn2_1/USPTO_spool/US09936377/runat_14112003_104450_22120/app_query.fasta_1.903
-DB=EST-QPMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936377 @CQN 1 1 2135 @runat_14112003_104450_22120 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	233.5	5.8	828	29	BZ565381	BZ565381 pacs2-164
2	185	4.6	983	29	BZ548740	BZ548740 pacs1-60
3	157	3.9	810	29	BZ551260	BZ551260 pacs1-60
4	153	3.8	960	29	BZ575443	BZ575443 msh2_4463
5	141	3.5	1218	29	BZ558672	BZ558672 pags8101_3
6	130.5	3.2	1526	29	BZ576993	BZ576993 msh2_520
7	122	3.0	911	29	BZ554877	BZ554877 pacs1-60
8	120.5	3.0	1354	29	BZ554885	BZ554885 pacs1-60
9	119.5	3.0	1161	29	BZ552855	BZ552855 pacs1-60
10	118.5	2.9	2124	11	AY105818	AY105818 Zea mays
11	117	2.9	897	29	BZ572021	BZ572021 msh2_233
12	117	2.9	2280	11	AK005069	AK005069 Mus muscu
13	115.5	2.9	585	10	BG456354	BG456354 NF077807P
14	114.5	2.8	765	13	BUI130232	BUI130232 603116821
15	112	2.8	610	28	BH374398	BH374398 AG-ND-153
16	112	2.8	1056	29	BZ548818	BZ548818 pacs1-60
17	112	2.8	1454	10	BF346294	BF346294 602018489
18	111.5	2.8	2341	11	AK031794	AK031794 Mus muscu
19	111	2.8	683	28	BH825061	BH825061 BACPP21-O
20	109.5	2.7	809	29	BZ597404	BZ597404 PUCCK35TD
21	109.5	2.7	1182	29	BZ565078	BZ565078 pacs2-164
22	109.5	2.7	1201	13	BX422599	BX422599 EX422599
23	109.5	2.7	1847	10	BF137846	BF137846 601782496
24	109	2.7	2864	11	AK043372	AK043372 Mus muscu
25	108.5	2.7	2169	11	AK016445	AK016445 Mus muscu
26	108	2.7	1120	13	BQ226095	BQ226095 AGENCOURT
27	108	2.7	5453	11	AK029845	AK029845 Mus muscu
28	107.5	2.7	837	13	BUI14889	BUI14889 603550605
29	107.5	2.7	960	13	BUS10302	BUS10302 AGENCOURT
30	107	2.7	262	10	BE123847	BE123847 BODAI10 BO
31	107	2.7	847	29	BZ548816	BZ548816 pacs1-60
32	107	2.7	874	29	BZ573965	BZ573965 msh2_3446
33	106.5	2.6	634	13	BH065834	BH065834 Fgr_8 K04
34	106.5	2.6	880	14	CA324304	CA324304 UI-M-FYO-
35	106	2.6	2531	11	AK016420	AK016420 Mus muscu
36	105.5	2.6	686	29	BZ564107	BZ564107 pacs2-164
37	105.5	2.6	1359	29	BZ563690	BZ563690 pacs2-164
38	105	2.6	858	10	BF969405	BF969405 602271677
39	105	2.6	858	13	BUI06765	BUI06765 603111672
40	104.5	2.6	932	9	AL969106	AL969106 AL969106
41	104.5	2.6	4015	11	AK036876	AK036876 Mus muscu
42	104	2.6	428	28	BH001150	BH001150 A2 Pirell
43	104	2.6	657	13	BUI064409	BUI064409 Fgr_4 M16
44	104	2.6	664	13	BUI060530	BUI060530 Fgr-C1-E
45	104	2.6	696	13	BUI065008	BUI065008 Fgr_6_G07

ALIGNMENTS

RESULT 1
BZ565381
LOCUS
DEFINITION
BZ565381 pacs2-164_522.sl pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION
BZ565381
VERSION
BZ565381.1
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 828)
828 bp DNA linear GSS 17-DEC-2002

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers

1. .828
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164-522"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 Whole genomic shotgun
library." 159 a 286 c 255 g 127 t 1 others

FEATURES

source

Alignment Scores:

Pred. No.: 1.56e-13 Length: 828
Score: 233.50 Matches: 77
Percent Similarity: 43.38% Conservative: 41
Best Local Similarity: 28.31% Mismatches: 114
Query Match: 5.79% Indels: 40
DB: 29 Gaps: 9

US-09-936-377-2 (1-758) x BZ565381 (1-828)

Qy 389 ProiledArgLeuLysGlySerTrpValGlnTyrLeuGlyGlnLysSerSerAla 408
Db 31 CCCTCGGACGCTGAGCGGGTGTGCGCGGAGTTCGCCACACCGCTTCTCGCC 90
Qy 409 LeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHis 428
Db 91 CTCGGC-----GAGGAAGCCTTCGTGCGCACACGAAACCGACAGG 132
Qy 429 TyrSerPheGlyValGlnAlaAsnTrp-----AspAsnPheThrLeuGlu 445
Db 133 GCCCGCTGTTCGCCCTCGAGGAA-----TGGAGCTCAGCGACCGCTCGACCTCAGC 186
Qy 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLleAsp 465
Db 187 TTCGGCGCCGCTGGAG-----CACACCGCGTGGACCCCGCGCGCAAGGC 234
Qy 466 ArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg 485
Db 235 AACGAGCGCTTCGCCGCAACGAC-----GGTTCGACAGAGCTTCACCCCGGC 282
Qy 486 SerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAla 505
Db 283 AGCTGTCCACCGCGCGGTGTACAGCTACGCGCGATCTGTCGTGCGCGCCACCTC 342
Qy 506 SerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAla 525
Db 343 AGCTACACCGAGCGCGCCCGACCTTCTACGAGCTGTACGCCACAGCTCGCGCGCGCC 402
Qy 526 ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu 545
Db 403 ACCGGCCTACGAGGTAGCGGTCGCGACGCGCAAGGAAAGGCGGTCTCCACCGAC 462
Qy 546 LeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArg 565
Db 463 CTCGCCCTGCGCTTCGACACCGCGGTGCACAAAGGCGCGGTGCGGGGTCTCTACGCGC 522
Qy 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLys----- 582

Db 523 TTCTCCAACATACATC-----GGGCTTCTCGCCAGCGGTCCCATCGCAACGAGGAAGGC 576
Qy 583 -----SerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 577 GAAGTGTGCGCGCGCGGATGACGAGGCGGTGCGGATACCTCTACAAAGCGGTTCG 636
Qy 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe-----LysProThrPro 615
Db 637 GCGGAATTCACGCGTTCGAGCCAGGCGCATTCACCTTCTGGAAGCCCGTACGCA 696
Qy 616 ArgTyrArgIleGlyValSerGlyAspTyrVal-----Arg 627
Db 697 ACTTTCGACTTGAACCTTTCGCGGGGATATACACCGCGCAAGAACAGGCAACGCGG 756
Qy 628 GlyArgLeu-LysAsnLeuProSerLeuProGly 638
Db 757 AACGGCTTGAAGGATTGCCCGCTTTCCTGGA 790

RESULT 2

BZ548740/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ548740 983 bp DNA linear GSS 17-DEC-2002
pacs1-60_1393.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
pacs1-60_1393, genomic survey sequence.
BZ548740
BZ548740.1 GI:27152321
GSS.
Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 983)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
Contact: Chris K. Raymond
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers

FEATURES
source

1. .983
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_1393"
/clone_lib="pacs1-60"
/notes="clinical isolate 1-60 Whole genomic shotgun
library." 153 a 298 c 321 g 211 t

BASE COUNT 153 a 298 c 321 g 211 t
ORIGIN

Alignment Scores:

Pred. No.: 2.14e-08 Length: 983
Score: 185.00 Matches: 89
Percent Similarity: 39.23% Conservative: 53
Best Local Similarity: 24.59% Mismatches: 148
Query Match: 4.58% Indels: 75
DB: 29 Gaps: 10

US-09-936-377-2 (1-758) x BZ548740 (1-983)

Qy 64 AspThrLeuArgGlnLysAlaVal-----AsnLeuGlyAspAlaLeuAsp 78
Db 975 GATTTCCTTAACGCAAGGTTTACCTTGGAAAAAAGGAACCTCGGAA-ACCCTTAAA 917
Qy 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArg 98

```

Db 916 GCTTTCCGGGTGCC-TTAACTATTGGGACGGGCCCAACCAACGGGTATCCGG 858
Qy 99 GylGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGly----- 115
Db 857 GGA-----TTGGAACGTTATCGGAATCCCGCTTTTCCAAA 822
Qy 116 -----AspMetAlaAspPheSerProAspHisAlaIleMetValAsp 129
Db 821 CGGGCTCGTGTGCGTTGAACGCGCTTGTGGTTGCTTAGACAAAGGGTTGCCGGAAGAC 762
Qy 130 -ThrAlaLeuSerGlnGlnValGluLeuArgGlyProValThrLeuLeuTyr-SerS 149
Db 761 CCCAAAAGTTGGAGCCCTCGAATGTACGCGCGCGCCCTTTGTTTACGGCTG 702
Qy 149 exGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProG 169
Db 701 GAAATGCAATCGCGGGGTGTTGAACAGCTCCGACCAACGATCCCAAGCGAC---CCGT 645
Qy 169 luAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeu 189
Db 644 TGGAGGCAATCAACGAGCAGCGGGAATCGCTACGCGCGCGCGCACACCCCGTAGCC 585
Qy 189 hrSerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu 209
Db 584 GTCCGGCGCACTGGAGCGCGCGACGCGC---AATTGCGCTGACGTGACGTGACCGCCA 528
Qy 209 yrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspS 229
Db 527 GCGCGGAGTTCAACGACGTC-----AGGATTCCTGTGA 495
Qy 229 exProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 249
Db 494 CGCCCATTCAGCGCGCGCGGAGATCGACGCGACAC---AAGC 448
Qy 249 yrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 269
Db 447 ATCGGTGCGAAGACAGCGCGCGCGGACGCGCGGACATCGGTGGCTCTATCACT 388
Qy 269 yrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuG 289
Db 387 GGGAGCAGCGTTACGCGCGCGCTCTCTAC----- 359
Qy 289 lnLeuTyrProHisLeuLeuThrGluAspValAspTyrAspAsnProGlyLeuSerC 309
Db 358 -----AGCGGTACGACGACGACATGCTGCTCGCGGCC----- 326
Qy 309 ysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleA 329
Db 325 -----GAGGACGAGTGGC-----CTGA 307
Qy 329 splLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349
Db 306 AGATGACGAGGACCGCTAGCCTTCGCTCCGAGATCCGCGACCTCGAAGGCGCGGTCA 247
Qy 349 luAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspA 369
Db 246 CTTGCTGAGCTGGAGCGCGCTATACCAAGTACGAGTACAGGAAATCAGAGATGGCG 187
Qy 369 laValGluAsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnP 389
Db 186 AGACCGGCGCACCATCTCAAGAACGAGGCTACGAAGGCGCGCATCGAGGCGCGCACCGCC 127
Qy 389 rolleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaL 409
Db 126 CGTCCGCGCGCTGAACGGGGTGGTGGCGCGCGAGTTTCGCCAACAGCGCGCTTCTCGGCC 67
Qy 409 eu 409
Db 66 TC 65

```

RESULT 3
BZ551260

```

LOCUS      BZ551260                      810 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pacsl-60_3063.y2 pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION  BZ551260
VERSION    BZ551260
KEYWORDS   BZ551260.1 GI:27154841
SOURCE     GSS.
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 810)
AUTHORS    Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE      Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL    J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 20622216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES   Location/Qualifiers
            source          1..810
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="1-60"
                        /db_xref="taxon:287"
                        /clone="pacsl-60_3063"
                        /clone_lib="pacsl-60"
                        /notes="clinical isolate 1-60 Whole genomic shotgun
                        library."
BASE COUNT 161 a 277 c 243 g 125 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1,29e-05      Length: 810
Score: 157.00           Matches: 65
Percent Similarity: 37.15% Conservative: 42
Best Local Similarity: 22.57% Mismatches: 110
Query Match: 3.89%      Indels: 72
DB: 29                  Gaps: 10
US-09-936-377-2 (1-758) x BZ551260 (1-810)
Qy 130 ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSer 149
Db 56 ACCCGGTGGCGCGCGCTCTAGAACTAGTAGTATGATGATCCCGCC----- 94
Qy 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu 169
Db 95 -----GGTTTGTGAACAGCTTCGACAAACCGCATCCCGACGAA---CCCGTC 139
Qy 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThr 189
Db 140 GATCGGATCCACGACGTCAGGATTCGCGGTACGCCATCCCGATCCCGAGCGCG 199
Qy 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyr 209
Db 200 TCCGGCGCATCGAGCGCGCGCGCGC---AATTGCGCTGACGTCGAGCGCGCGC 256
Qy 210 ArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSer 229
Db 257 CGCGAGTTCAAACGACGTCAGGATTCGCGGTACGCCATCCCGATCCCGAGCGCG----- 304
Qy 230 ProArgArgPhe-----AlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArg 247
Db 305 CAGCGCAGATCGACCGCGCACCGCGCAAGCATCGGGTG----- 343
Qy 248 PheTyrArgArgThrTyrSerAspArgArgGlnTyrGlyLeuProAlaHisSerHis 267

```


[illegible]

Alignment Scores:

JOURNAL
COMMENT

Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source

Location/Qualifiers
1..1354
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_491"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 292 a 390 c 381 g 288 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 0.165 Length: 1354
Score: 120.50 Matches: 65
Percent Similarity: 36.10% Conservative: 35
Best Local Similarity: 23.47% Mismatches: 83
Query Match: 2.99% Indels: 94
DB: 29 Gaps: 14

US-09-936-377-2 (1-758) x BZ554885 (1-1354)

QY 39 ValGlyLysSerArgProAlaThrSerGlyLeuLeuHisThrSerThraAlaSerAsp 58
Db 715 GTTGGCGAAGCGGA-----CACAAAGTTGTGTAATAAC 680

QY 59 Lysile-----lleSerGlyAspThrLeuArgGlnLys-AlaValas 72
Db 679 CATCTCGGTGCCGCAACTGGTCGTGTCGGAATGCTCGAGAGCGGTGCAGAA 620

QY 72 nLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrglyGlyAl 92
Db 619 CGTCCGCGCAGGTCCTCGCAAGTGCGCGCGTCGCAAGTGCAGGACAACACCGTACC 560

QY 92 aSerAlaProVal-----lleArgGlyGlnThrGlyArgAlleLysVa 107
Db 559 CGCAGCGCATCTCCTCGAACGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511

QY 107 lLeuAenHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlalleMe 127
Db 510 -----CTGTGCGCGCTCGACGGTGAT 488

QY 127 tValasp-----ThrAl 131
Db 487 GATCGACGGCGTCCGCGCGCGTGGCGCTTACGCGCAGCGCGAGCTGTGATGATGCC 428

QY 131 aLeuSer-----GlnGlnValGluilleLeuargGlyProValThrLeuTy 147
Db 427 GCTGTCCATCGGCAACTCGAGAGCATCATCGTGTGCGCGCGCGCTGTCGCGCTA 368

QY 147 rSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMe 167
Db 367 CGGCCCCGCAAGCTCGCGGGGTGATCACTTCGTGACCCGGCGATTCCGAGAAGTT 308

QY 167 tProGluAenGlyValSerGlyGluLeuGlyLeuargLeu-----SerSerGlyAs 184
Db 307 C-----TCCGGGAATATCGCACCAACCATCGAGCATCGCGCCACGCGCG 263

QY 184 nLeuGlyLysLeuThr-----SerGlyGlylleAenlleGlyLeuGlyLysAs 200
Db 262 CTGAGAGAGCTCAACCGCGGTCTCTCGCGGGGACCGCGCACACCGCGCTGGGC----- 208

QY 200 nPheValLeuHisThrGluGlyLeuTyArgLysSerGlyAspTyAlaValProArgTy 220
Db 207 -GTGGCGCTGTGTATTCGGGGTG-----AAGGGCGCGACTATC----- 169

QY 220 rArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAl 240
Db 168 -----CGCGACGGCACACAGCAATGATAT 143

QY 240 aValLeuGlyTrpArgLysArgPheTyArgArgThrTySerAspArgArgAspGlnTy 260
Db 142 CGACGACGTGCTGCTCAGACCCATTGGCAGCTCACCGACAGCGACCGAG----- 94

QY 260 rGlyLeuProAlaHisSerHisGluTyAspAspCysHisAlaAspIle 276
Db 93 ----TTGGCGGCGCAACTTCCACTACTAC---GACCGCTACGCCGATATG 52

RESULT 9
BZ552855/c 1161 bp DNA linear GSS 17-DEC-2000
LOCUS pacsl-60_3853.y2 pacsl-60 Pseudomonas aeruginosa genomic clone
DEFINITION pacsl-60_3853, genomic survey sequence.
ACCESSION BZ552855
VERSION BZ552855.1 GI:27158033
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequencing variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source

Location/Qualifiers
1..1161
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_3853"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 194 a 351 c 312 g 302 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 0.168 Length: 1161
Score: 119.50 Matches: 44
Percent Similarity: 37.06% Conservative: 29
Best Local Similarity: 22.34% Mismatches: 78
Query Match: 2.96% Indels: 46
DB: 29 Gaps: 5

US-09-936-377-2 (1-758) x BZ552855 (1-1161)

QY 218 ProArgTyArgAsnLeuLysArgLeuProAspSerPro-----ArgArgPhe 233
Db 623 CCSCGAGTTCAACAGACGTCCAGGATTCCCGCTACGCCACATTCACGCCGCGCGAG 564

QY 234 AlaAsnGlyClHiHiaArgAlaValLeuGlyTrpArgLysArgPheTyArgArgThrTy 253
Db 563 ATCGAGCGCGACCGCGNCAAGCATCGGTGTCAGAACAGCA----- 522


```

Db      555 AGTTCGTTGTCGGCCTGGCGGTCGCATCCGCTGGTGTGTCATGACGAGGCGGCAG 496
Qy      24  AlahisclutrgluInserValGlyLeuGluThrValThrValValGlyLysSerArg 43
Db      495 GCCGAAGAGGAGCAAGGAGCTGGGC-----ACGGTCACCGTGTGCGGACTGGCTG 442
Qy      44  ProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle---Ser 62
Db      441 GGCAGAGCCGACACAGCGGTGGTGGCAGAACCATCCCGGTGCGCGACGCTGGTGGTGGT 382
Qy      63  GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db      381 CGGGAATGCTCAGAGCGGTGGCAGAACGTCGCGACGT-GTGGCAAGGTGCCCGC 323
Qy      83  IleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal-----Ile 97
Db      322 GTCCAAGTCGAGGACCAACACCGGTACCGGCGGACGACATCTCCCTGAACGTGGGGTG 263
Qy      98  ArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMet 117
Db      262 CGTGGCTGACTTCACGC----- 245
Qy      118 AlaAspPheSerProAspHisAlaIleMetValAsp----- 129
Db      244 -----CTGTCGCGCGCTCGACGGTGTATCGACGGCTGCCGGCGGTGGCGCC 191
Qy      130 -----ThrAlaLeuSer-----GlnGlnValGlu 137
Db      190 TAGCGGCAGCGCAGCTGTCGATGATCCGCTGTCCATCGCAACCTGGAGAGCATCGAC 131
Qy      138 IleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAsp 157
Db      130 GTGTCGCGCGCGCGCTCGGTGCGCTACGCGCGCAGAACGTCGCGGGGTGATCAAC 71
Qy      158 ValAlaAspGlyLysIleProGlyLysMetProGluAsnGlyValSerGlyGluLeuGly 177
Db      70  TTCGTGACCCGGCGATTCGGAGAGTTC---TCCGGCGGGGGATCCTCTAGAGTCGAC 14
Qy      178 LeuArgLeu 180
Db      13  CTGAGGCTA 5

RESULT 17
BF346294
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
BF346294
VERSION
BF346294.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1454)
NTH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LIA9423
High quality sequence stop: 62.
Location/Qualifiers
1..1454
/organism="Homo sapiens"

```

```

/molecule_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4154014"
/tissue_type="anaplastic oligodendroglioma with lp/19q
loss"
/Tab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Brn67"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 609 a 502 c 305 g 38 t
ORIGIN
Alignment Scores:
Pred. No.: 1.39 Length: 1454
Score: 112.00 Matches: 88
Percent Similarity: 33.09% Conservative: 50
Best Local Similarity: 21.10% Mismatches: 168
Query Match: 2.78% Indels: 111
DB: 10 Gaps: 20
US-09-936-377-2 (1-758) x BF346294 (1-1454)
Qy 204 HistThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeu 223
Db 193 CATCAAGAGGAGCAAGAGAGCAGCAGCGAG-----ACAGAG 231
Qy 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db 232 AAGCGACACCGGACACGACAGAG-----GGCGCATCACACGAGCGCGCGAAGCTCAAC 285
Qy 244 TrpArgLysArgPheTyrArgArgThrTyrSerAspArg----- 256
Db 286 ACCAGGCAAGAGCAGCAGCAAAACAAACCAACCAATAGCATAGAAAGAACCCCGAG 345
Qy 257 -----ArgAspGlnTyrGlyLeuProAlaHisSerHisGlu-----Tyr 269
Db 346 GAACCGGAAAGACACGCAACAAACACGACGACCAACACACGACGAGCGCGCAGGA 405
Qy 270 AspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLys-----ArgTyr 287
Db 406 AACGACGCGCGCAGATCTCACAACCGCGACACACAGACGACCAAAACAGACAGCGAC 465
Qy 288 LeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeu 307
Db 466 CGCAAGAAAGCCCGCCAC-----ACGAAGCCCAACGACGACCAAGAC----- 507
Qy 308 SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysPro--- 326
Db 508 ---GCCGGAACAAACAGAGAC-----GCACGCCACACGACGACCAACCCCA 549
Qy 327 -----TrpIleAspLeuArgAsnLysArg 334
Db 550 CAGACAAAGACACGCGCCAAAGACCAAGCCACCATGAGTGGACACACAGACGCGCAAGAGA 609
Qy 335 TyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
Db 610 -----ACACAAGCCGACACACGACGACCGCGACCC-----AGAAACGAAACCGCGAC 654
Qy 355 LeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGlu-AsnPhePh 374
Db 655 CAGACGACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 709
Qy 374 eAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProfile----- 390
Db 710 -AACACGAAACCGGAAACCGGACACACACACACACACGAAAGAGACACTCACAGCCCAACA 768
Qy 391 -----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGln----- 404
Db 769 CGACGGCACACAACTAGAGGACGAAACGCGGACACACACGCGGCAACGAGCCCAAGCCCC 828
Qy 405 -----LysSe 406

```

```

Db      829 AGCCCGCCAGCAGACACAAAGACAAAGACACACACAGAGCGCGCACATACCGGC 888
      Qy      : rSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspHisLysVa 426
Db      889 ACGCGAAACAGGTAAGCAAAAGAGCAAAAGAGAGCGCCCGGACAAACCAACACACACA 948
      Qy      : lGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluG1 446
Db      949 ACAACACCGG-----GAACCGCACACAGGACAAACCCCGCACAAACACACAG 993
      Qy      : yGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspAr 466
Db      994 ACCAAGCACACACAGGCGGACACACAGGCAACACACAAAGAGAGGGAAGCGCGAGCG 1053
      Qy      : gGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln--ThrAlaAr 485
Db      1054 CGCACACACACACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1113
      Qy      : gSerPheAlaLeuSerGlyAsnTrp---TyrPheThrProGln-----HisLysLeuSe 502
Db      1114 AACCAGAGCGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1173
      Qy      : rLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db      1174 CCAGAACGAAAGCCACACAGCACTCCCGACACACC---CCCGAGCGAGCGACACACAAA 1230
      Qy      : s-----HisValAlaThrAs 527
Db      1231 ACGAACAGACACACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1290
      Qy      : nThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db      1291 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1339

RESULT 18
AK031794
LOCUS
DEFINITION Mus musculus 11 days embryo head cDNA, RIKEN full-length enriched
            library, clone:6230428F13 product:X-linked myotubular myopathy gene
            1. full insert sequence.
VERSION AK031794 2341 bp mRNA linear HTC 05-DEC-2002
KEYWORDS Mus musculus 11 days embryo head cDNA, RIKEN full-length enriched
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, N., Nishine, T., Harada, A.,
Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipipillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

```

```

20530913
11076861
4
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Harai, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staehli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustigich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
and Hayashizaki, Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2341)
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
Location/Qualifiers
1. 2341
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:6230428F13"
/db_xref="taxon:10090"
/clone="6230428F13"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="11 days embryo"
5. 1124
CDS

```



```

Score: 111.00 Matches: 45
Percent Similarity: 43.64% Conservative: 27
Best Local Similarity: 27.27% Mismatches: 61
Query Match: 2.75% Indels: 32
DB: 28 Gaps: 8

US-09-936-377-2 (1-758) x BH825061 (1-683)

Qy 29 GlnSerValGlyLeuGluThrValThrValValGlyLys-----SerArgProArgAla 46
Db 45 CAGGATACCGCCCGGATACCTCTCGTGTACTGTCTAAACCGTTTGAACAGCGCGGAGC 104
Qy 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleSerGlyAspThrLeu 66
Db 105 ACT-----GTGCTTGCACCAACCAACCGTT-----GTGACCCGTCAGGATATCGAC 149
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyLysIleHisAlaSer 86
Db 150 CGCTGGTAGTGCACCTCGGTCAATGATGTGCTGCCCGCTCTCCGGGCTCGATATCAC 209
Qy 87 GlnTyrGlyGlyAla-----SerAlaProValIleArgGlyGlnThrGlyArg 104
Db 210 CAAAACCGCGGTTCAGGTTCAGCTCTCATCTATTTTATTCGGCGGTACAAATGCCAGTCAT 269
Qy 105 IleLysVal-----LeuAsnHisHisGlyGlyGluThrGlyAspMetAla 118
Db 270 GTGTGGTGTAAATGATGGCGTACGCTGAATCTGGCGGGGGTTCAGTGTCT---GCC 326
Qy 119 AspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIle 138
Db 327 GACCTTAGC-----CAGTTCCCTATTGGCTTGTCCAGCGTGTGAATAT 371
Qy 139 LeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal 158
Db 372 ATCCCGTGGCGCGCTCCGCTCTTTATGTTCCGATGCAATAGCGGGGTGTGAATATC 431
Qy 159 AlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeu 178
Db 432 ATCAGACGCGC-----GATGACCCCGAAGC 458

Qy 179 ArgLeuSerSerGly 183
Db 459 GAAATTCAGCAGG 473

RESULT 20
BZ697404/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BZ697404 809 bp DNA linear GSS 19-FEB-2003
PUCBK35TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTal27E21,
genomic survey sequence.
BZ697404
BZ697404.1 GI:28417251
GSS.
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1. (bases 1 to 809)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick
, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.
Maize Genomics Consortium
Unpublished
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1. .809
/organism="Zea mays"

FEATURES
source
BZ6565078
pacs2-164_5068.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_5068, genomic survey sequence.

/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTal27E21"
/clone_lib="ZM 0.6 1.0 KB"
/note="Vector: pCR4-TOPO, Site 1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

BASE COUNT 111 a 286 c 250 g 162 t
ORIGIN
Alignment Scores:
Pred. No.: 1.09 Length: 809
Score: 109.50 Matches: 58
Percent Similarity: 34.78% Conservative: 22
Best Local Similarity: 25.22% Mismatches: 65
Query Match: 2.71% Indels: 85
DB: 29 Gaps: 10

US-09-936-377-2 (1-758) x BZ697404 (1-809)
Qy 471 LysGlnProLeuProAspLeuGlyAlaHisArgGlnThr----- 483
Db 796 AAGCAGGGGCTGCCGCGAACCGCGCGCGCAGCACGTGGTGGCGACATCAAGTTGT 737
Qy 484 AlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db 736 TCTCGTTCAGCGTCACGCGCGCGCGCGACATTCGGCCCTTAAGACCCATTGAGCGAC 677
Qy 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db 676 CGAGTTCCCTCGGACCGCGACCCAGTCAGTCGGGACTACCGCTGAGTTTAAAGCAT 617
Qy 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db 616 ATA-----ATAAGCGAGGAGGAAGAAC 593
Qy 544 IleGluLeuAlaLeuGlyTyr-----GluGlyAspArgTrpGlnTyrAsnLeu 559
Db 592 TTACGA-----GGATTCCCTTAGTAACGCGCGAGCAACCGGAGCAGCCGACTTG 542
Qy 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArg 579
Db 541 AGAATCGGGCGGCTCGCGCGCGGANTTGTATCTGGAGAGGCGTCTTAGCGACGACCG 482
Qy 580 GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 481 GCGCCCAAGTCTCTCGAAAGGACGCC----- 455
Qy 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro----- 615
Db 454 -----TGGAGGGTGAGAGACCCCGCTCCGCGCGGACCTGTGCGACCCAGC 410
Qy 616 -----ArgTyrArgIleGlyValSerGlyAspTyrVal----- 626
Db 409 AGGCCCGCTCAACGAGTCGGGTGTTTGGGAATGACGCCCAATCGGGCGGTAAACTCGC 350
Qy 627 -----ArgGly----- 628
Db 349 TCCAAGGCTAAATACAGCGGAGAGACCGATAGCAACAAAGTACCGCGAGGAAGATGAA 290
Qy 629 -----ArgLeuLysAsnLeuProSerLeuProGlyArgGlu----- 640
Db 289 AAGGACTTTGAAAGAGAGTCAAGAGTGTGCTTGAATTCGCCGGGGAAGCGGATGGGG 230
Qy 641 -----AspAlaTyrGlyAsnArgProPhe 648
Db 229 GCTGCGAGCGGCAC-----CGGCCGTAT 206

RESULT 21
BZ565078/c
LOCUS
DEFINITION

```


Qy 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGlyLeuMetProGlu 169
 Db 270 GGCCTCAATGAAGGAGAGATTACATCAAAATATCTCTTTATTAAGAAGTTTGGA 329
 Qy 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSer---SerGlyAsnLeuGlyLysLeu 188
 Db 330 ACGGATCTTCTTAATATCTGATGCTCTCTGGGTGTGATCTCGAGAAATGAAAAATG 389
 Qy 189 -----ThrSerGlyGlyLeuAsn---IleGlyLeu----- 197
 Db 390 GGAGCGCGGACAGTAGAGAGAAATTCCTATGCTAGATATTAATCTGTAAGACATG 449
 Qy 198 ---GlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSer----- 212
 Db 450 AGAAACCTGAGGTTGCTTTGAAACAGGAAGGCCACAGCAGAGAGATATGTTGAGATC 509
 Qy 213 ---GlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArg 231
 Db 510 CTCAGAGATACGGCTTTCCCTCGCTCACAGTCTGCCATTAATTT----- 554
 Qy 232 ArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArg 251
 Db 555 GCATTTTAAATGAAGAAAGTTTAACGTGGATGATGG-----ACAGTTTACAATCCA 608
 Qy 252 ThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAsp 271
 Db 609 GTGGAAGAATACAGGAGCGAG-----GCCTTGCCCAATCACCAT----- 647
 Qy 272 CysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr-----LeuGln 289
 Db 648 -----TGGAATAACTTTTATTAATAAGTGTGCTATGAGTCTGTGAC 689
 Qy 290 LeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys 309
 Db 690 ACTTACCCTGCTCTTTGCTG-----GTTCCGTATCGT----- 722
 Qy 310 GlyPheHisAspAspAsp-----AlaHisAlaHisAlaHisAsnGlyLys 325
 Db 723 -----GCCTCAGATGATGACCTCCGAGAGATTGCAACTTTTGTGCTCCGAAATCGAATT 776
 Qy 326 Pro-----TyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
 Db 777 CCAGTGTGTATGATGATTCATCCAGAAATAGAGCGTCATGTGCGTGC-----AGT 830
 Qy 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHis 362
 Db 831 CAGCCTCTTGTGCGTATGAGTGGGAAACGA-----ATAAAGAT----- 869
 Qy 363 AspGlyLysAlaGlyAspAlaVal-----GluAsnPhe 373
 Db 870 GATGAGAAATATCTCGATGTTATCAGGGAGACTAATAACAAATTTTC 917
 RESULT 23
 LOCUS BF137846 1847 bp mRNA linear EST 24-OCT-2000
 DEFINITION 601782496F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4010734 5',
 mRNA sequence.
 ACCESSION BF137846
 VERSION BF137846.1 GI:10976886
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1847)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LAM9248 row: o column: 23
 High quality sequence stop: 541.
 Location/Qualifiers

FEATURES

source

1. 1847
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CZECH II"
 /db_xref="taxon:10090"
 /clone="IMAGE:4010734"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Lu30"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; transgenic model WNT-1, expression driven by
 MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
 dt. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 661 a 446 c 546 g 193 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 3.57 Length: 1847
 Score: 109.50 Matches: 121
 Percent Similarity: 31.73% Conservative: 50
 Best Local Similarity: 22.45% Mismatches: 191
 Query Match: 2.71% Indels: 179
 DB: 10 Gaps: 26

US-09-936-377-2 (1-758) x BF137846 (1-1847)

Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
 Db 519 GTAGNCAATGATCATGAAGAAGAT-----CTTGACTCGCGAAGAACTCATCGGC 569
 Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu 146
 Db 570 CAGGTGGAAGACAGATTGGCCAGCA-----CTG 599
 Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGlyLys 166
 Db 600 GTAGCAGTGTGCCACATGCTGGTTGGTTCCCTA-----AAA 638
 Qy 167 MetProGluAsnGlyValSerGlyGlu---LeuGlyLeuArgLeuSerSerGlyAsnLeu 185
 Db 639 CTAAAGGAGGTGAGTGGCGGAGACAGCGGTGAGACGCTCGCGGGGGGCAAC--- 695
 Qy 186 GluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThr 205
 Db 696 -----ACTGCAAGTGCCTCAACCGCGGGGGGCGGACAAAT----- 731
 Qy 206 GluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225
 Db 731 ----- 731
 Qy 226 LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArg 245
 Db 732 ---GATCCACAGCCCAAGCGG---ACCGGGGACACACCGTGGCACCTCAGCGGGCGT 785
 Qy 246 LysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHis 265
 Db 786 AACCAAC-----CTTCGGTTTCC 803
 Qy 266 SerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuLeuLys 285
 Db 804 TCACACCATAGTAGC-----CATCGG-----TGGTGGCTATCATGCTACGCAGA 848
 Qy 286 ArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnPro 305

Db	849	CGCACCAAGCACCACAGCCACATATA-----	875
Qy	306	GlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLys	325
Db	876	-----CGAAA-CACACACGAAGA	892
Qy	326	ProTrpIleAspLeuArgAsnLys-----ArgTyrGluLeuArgAlaGluTrpLys	342
Db	893	CCAAGCATAGATCACAGAGAGAAACACAGACAGACTAAATACGACACAGACCCACAAA	952
Qy	343	--GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis	362
Db	953	CACACCA-----	CGAC 964
Qy	362	isAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGln-----	379
Db	965	ACAACGACCCAGCAAAAGAAACAAACACAGAACACACAGCAACAAACAGCAACAAACAAA	1024
Qy	380	-----AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlyLys	396
Db	1025	CAAAACACACCAACGACCGGGAG--CGGAGGGCAAGAGCGGGAGGAACGGGGAA	1081
Qy	396	erTrpGlyValGln-----TyrLeuGlyGlnLysSerS	407
Db	1082	CGCGAGCGCACAGACAGAGGCACAAAGGGAAGAGACACAGCAGAGGAGAGAGGCAC	1141
Qy	407	erAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValG	427
Db	1142	AGCAGCGGACGACCGGCAAGACCGGA--AAGAAGAAACGAGAGACGAA--AAGAGCG	1197
Qy	427	lnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGlu---	445
Db	1198	AGCATGAGAGACCAAGGAGGAGAGAGAGCGGCAAGAGAC-----ACAGAGAAAGAGA	1251
Qy	446	-----GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaL	463
Db	1252	AGAGGAACGGAGGAGCGGCAGACGACGAGCAGCAGCAGAGAGAGAGAGCAGGA-----	1304
Qy	463	eulleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnT	483
Db	1305	-----GACCGAGGAGAGCGGAGAGAGAACACAGAGAGAGAC-----CCAG	1344
Qy	483	hrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerL	503
Db	1345	CGCAGAGAGAGAGAGCGCAGGAGGAGAG-----	1373
Qy	503	euthrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysH	523
Db	1374	-----ACGCACGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGC	1425
Qy	523	isValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsn-----	537
Db	1426	ACCAGCCGACAGACAGACGCA--GGAAACAAAGAGAGAGAGAGAGAGAGAGAGAG	1482
Qy	538	-----LysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyA	553
Db	1483	GCAGAGAAACCCAGAGAGACGACGACGACGACGACGACGACGACGACGACGACGAG	1542
Qy	553	spArgTTPGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaG	573
Db	1543	GAACAGCAGCAGCAGCAGCAGC-----AGCACACACACACACAGAGAGAGAGAGCC	1596
Qy	573	lnThrLeuAsnAspGlyArgGlyProLysSerIleGlu-----AspAspSerGluMetL	591
Db	1597	AACGAGGAGGTGACAAACACAGCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1656
Qy	591	ysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu	608
Db	1657	AGGCGAGCAGCAGAGAACAGCGCGCAACCGCAACCGCAACCGCAACCGCAACCGG	1709
Db	1484	AK043372	2864 bp mRNA linear HTC 05-DEC-2002

RESULT 24
AK043372
LOCUS

DEFINITION	Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730089G18 product:hypothetical PWWP domain containing protein, full insert sequence.
ACCESSION	AK043372
VERSION	AK043372.1 GI:26335654
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636
AUTHORS	2
TITLE	3
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, O., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL	4
MEDLINE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldairelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851
REFERENCE	5
AUTHORS	Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2864)
TITLE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730089G18 product:hypothetical PWWP domain containing protein, full insert sequence.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, O., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldairelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2864)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,


```

US-09-936-377-2 (1-758) x AK016445 (1-2169)
QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaaspPheSerProAspHis 124
DB 250 AITAGAAGTGTCTCAAGATGAGTCACTGAGTGTGAGGAGACTGCTCCCTCGG--- 306
QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu----- 139
DB 307 -----CTCCAGGGGAGTTACTAATTACTGAAAAGAAAGTTATTATACATATGTCCTTC 360
QY 140 ArgGlyProValThrLeuLeuTyrSerSerGlyAsnVal----- 152
DB 361 AATGGCCCCATT-----AAGGGAGAGTTTACATCAAAATTATCGTCTT 405
QY 153 -----AlaGlyLeuValAspValAlaAspGlyLysIle 163
DB 406 TATTTAAGAAAGTTTGAAGAGGATTCGTCTTAATACTTGAATGTTCTCTGGGTGATA 465
QY 164 Pro-----GluLysMetProGluAsnGlyValSerGlyGlu-----LeuGlyLeuArg 179
DB 466 TCAAGAAATTGAAAAATGGAGGCGCACAAAGTAGAGGAGAAAATTCCTATGCTTAGAT 525
QY 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLys 199
DB 526 ATTACTTGTAAAGATTGAGAACCTG----- 552
QY 200 AsnPheValLeuHisThrGluGlyLeuTyrArgLysSer-----GlyAsp 214
DB 553 AGGTTTGCAATTGAAGCAAGAGGCCACAGCAGAGAGATATGTTTGAGATCCTTGTAAAA 612
QY 215 TyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgPheAla 234
DB 613 CATGCTCTTCTCTGCGCACACAATCTGCCATTATTT-----GCATTTGTA 657
QY 235 AsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSer 254
DB 658 AATGAAGAGAAGTTTAACGTGGATGGGTGG-----ACTGTTTATATCAGCTTGAGAA 711
QY 255 AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspCysHisAla 274
DB 712 TATAGAAGGCAG-----GSCCTGCCCAATCACCAT----- 741
QY 275 AspIleIleTrpGlnLysSerLeuIleAsnLysArgTyr-----LeuGlnLeuTyrPro 292
DB 742 -----TGGAGGATAAGTTTATTAACAAGTGTCTATGAGTCTCTGAGACATACCCT 792
QY 293 HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHis 312
DB 793 GCTCTTTTGGTG-----GTTCCCTATCGACC----- 819
QY 313 AspAspAspAsp-----AlaHisAlaHisAlaHisAsnGlyLysPro----- 326
DB 820 TCAGATGATGATCTTAGGAGGATCGCAACGTTTAGATCCGAAATCGGCTTCTCTGACTG 879
QY 327 ---TrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPhe 345
DB 880 TCGTGATTCACCCAGAAACAAATGCTCAITATGCGCTGC-----ACTGAGCTCTT 933
QY 346 ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLys 365
DB 934 GTCGGTATGATGGTAAAGA-----AATAAGAT-----GACGAGAA 972
QY 366 AlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeu 385
DB 973 TACCTGGATGTGATCAGGGAA-----ACTAACAAACAACTTCTAAGCTCATGATTATGAT 1029
QY 386 ArgHisGlnProIleGlyArgLeu-----LysGlySerTrpGly----- 398
DB 1030 GCACGACCCAGTGAATGCGTCGCCAACAGAGCAGGAGGAGATATGAAAGTAT 1089
QY 399 ValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGln 418

```

```

DB 1090 GACGCATATCAAAACTCAGAACTTCCTTCTTAGACATTCATAATATTCATGTTATGCCGA 1149
QY 419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
DB 1150 GAATCTTTA-----AAAAAGTGAAGATATGTTATCCCAACATAGAGATCTCAT 1203
QY 439 Trp-----AspAsnPheThrLeuGluGly 446
DB 1204 TGGTGTCTCCAGTTTGAGTCTACTCATTTGTTTAGAACATATCAAGCTGTGTTCTGACCGGT 1263
QY 447 GlyValArgValGluLysGln-----LysAlaSerIle----- 457
DB 1264 GCATTCAGTGCAGACCAAGTGTCTTCAGAAAGAGCTCGGTATCTTGTGCACTGCAGT 1323
QY 458 ---ArgTyrAspLys-----AlaLeuIleAspArgGluAsnTyr 469
DB 1324 GACGGATGGGACAGGACCCGCTCAGCTGACATCCTTGGCCATGCTGATGTTGGACAGCTTC 1383
QY 470 TyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
DB 1384 TACAGA-----ACTATTGAAGCTTTGAGATA 1410
QY 490 -----SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis 507
DB 1411 TTGCTACAGAAAGAGTGG---ATAAGTTTTGGCCATAAATTTGCATCTAGATAGTCTCAT 1467
QY 508 GlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn 527
DB 1468 GGTGATAAA----- 1476
QY 528 ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAla 547
DB 1477 -----AACCATGCTGATGCTGATCGATCT--- 1509
QY 548 LeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAla----- 560
DB 1510 CTTCAGTTTATGACTGTGTGTGGCAGATGTGCAACAGTTCGCCACAGCTTTTGAGTTC 1569
QY 561 -----LeuTyrArgAsnArgPheGlyAsnTyr 569
DB 1570 AATGAAGGCTTTTGTATTACCGTTTGGATCATCTGTATAGCTGTGCTGATTTGTTACTTTC 1629
QY 570 IleTyr 571
DB 1630 TTATTC 1635

```

```

RESULT 26
BO226095
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
BO226095
VERSION
BO226095.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1120)
AUTHORS
NIH-MGC http://mgs.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
Found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13346 row: i column: 24
High quality sequence stop: 470.

```

FEATURES
source

Location/Qualifiers
1..1120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6067295"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library." 1 others

BASE COUNT 254 a 354 c 294 g 217 t

ORIGIN

Alignment Scores:
Pred. No.: 2.48 Length: 1120
Score: 108.00 Matches: 75
Percent Similarity: 34.44% Conservative: 29
Best Local Similarity: 24.83% Mismatches: 119
Query Match: 2.68% Indels: 81
DB: 13 Gaps: 14

US-09-936-377-2 (1-758) x BQ226095 (1-1120)

QY 10 ValLeuSerIleLeuLeuLeuAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGln 29
:: ||| :: ||| :: |
Db 90 CTTTCTCTACTGTGTCACTAAACAGCGTGTTATCATGAGCAGCAACTGGCTTCGTGA 149
:: ||| :: ||| :: |
QY 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThr---Ser 48
:: ||| :: ||| :: |
Db 150 GC AACCGA AATTCACGACGAGAACGTTCAAAGTGTGACGCCGAGGTGCGAGGCTCCCTCT 209
:: ||| :: ||| :: |
QY 49 GlyLeuLeuHisThrSerThrAlaSerAspLysIlelleSerGlyAspThrLeuArg--- 67
::::|||::::|||::::|||::::|||::::|||::::|||::::|||::::|||::::|||::::|||
Db 210 AGAGTGATCCATCGGNAAGTCCCCATCGACGTGACGGAGGGGAGTAGTCATCTCCCTG 269
:: ||| :: ||| :: |
QY 68 -----GlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIle 83
:: ||| :: ||| :: |
Db 270 GGSGCTCCCTTTGGGAAGTACCACACTCTCT-GATGCTGAAGGGGAAAAACAGGCTT 328
:: ||| :: ||| :: |
QY 84 HisAlaSerGlnTyRGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArg 103
:: ||| :: ||| :: |
Db 329 CATCGAGATGAACAGGAGGAGGCTGC----- 355
:: ||| :: ||| :: |
QY 104 ArgIleLysValLeuAsnHisHisGlyGlu-----ThrGlyAspMetAlaAspPhe 120
:: ||| :: ||| :: |
Db 356 -----CAACACCATGGTGAATACTACTACCTCGGTGACCCCTGTGCTGGCG 400
:: ||| :: ||| :: |
QY 121 SerProAspHisAlaIleMetVal-----AspThrAlaLeuSer 133
:: ||| :: ||| :: |
Db 401 CGGCCAGCCCATCTACATCCAGTTCTCCAACCAAGGAGCTGAAGCCGACGACTCTCC 460
:: ||| :: ||| :: |
QY 134 GlnGlnValGluLeuArgGlyProValThrLeuLeuTySer----- 148
:: ||| :: ||| :: |
Db 461 CAACCAAG-----CGCGGGCCCAAGCGGGCCCTCGACGGCGGTAACCTGGCTCCAG 510
:: ||| :: ||| :: |
QY 149 SerGlyAsnValAla-----GlyLeuValAspValAlaAspGlyLys 162
:: ||| :: ||| :: |
Db 511 TCGGGGAACCTGGCTTGGCTGCTCGCGCGCGCGCTGGACCCAGGGGATGGCGA 570
:: ||| :: ||| :: |
QY 163 Ile-ProGlyLysMetProGluAsnGlyValserGlyGluLeuGlyLeuArgLeuSerSe 182
:: ||| :: ||| :: |
Db 571 ATGGCCC-----GGGCCCAAGAGCCCCCGTGCCTTCGCGAATCCATC 612
:: ||| :: ||| :: |
QY 182 rGlyAsnLeuGluLysLeuThrSerGly-----GlyleasnlleGlyLeuGlyLy 199
:: ||| :: ||| :: |
Db 613 CGGGGGAAAAAACCCTCTTTCAACCTCTCGTGGGAAACCCCTTTGGATTG---- 667
:: ||| :: ||| :: |
QY 199 sAnPheValLeuHisThrGluGlyLeuTyRArgLysSerGlyAspTyRAlaValProAr 219
:: ||| :: ||| :: |

[illegible]


```

Db      2836 CGAGATAGAAATCCTAGCCATCAATACGACAGCTTGAAGGGAAGCCTCTGAGTGAAGC 2895
Qy      343 GlnProPhe---ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
Db      2896 CATCCACTTCTCCAGAT-----GGCAGAGAGAGACTGTGCAC 2931
Qy      362 HisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAla 381
Db      2932 CCTGAAATTAAGAAACAGACAGATGGCTCACTTT-----TCAGATTC 2976
Qy      382 ArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyr 401
Db      2977 AGATACAAATACACACCT-ATCATTTGA-----GGATCCAA--- 3014
Qy      402 LeuGlyGlnLysSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeu 421
Db      3015 -----AGCAAGA-ACCAGCTGTCC-----CCAGTCCCAAGCCT----- 3049
Qy      422 LeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyrAspAsn 441
Db      3050 -----CGAAGCCAGACGTACCCAGATGTGGCCTGAGTAATGAAGACTGGGATCGA 3100
Qy      442 PheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLys 461
Db      3101 TCCACAGCCAGTGTCTTTAGGGCTCT-----GACAGT 3136
Qy      462 AlaLeuIleAspArg-----GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
Db      3137 GCAGATCTGAACAAGAGAAACTTCTGTCTCAAGCATTTGGAGACCTGGAGACCTGC 3196
Qy      481 ArgGlnThrAla-----ArgSerPheAlaLeuSerGlyAsnTyrPhe 495
Db      3197 GGCAGTGGGGATCTCGAGACCTGAGCAACAATCTGCGGGAGT---ACTATG 3253
Qy      496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGln 515
Db      3254 AGTTTGAATCATGAGCTCCAACTCGCAGTCAGCTGGGCGACAGCGCAGCTTCCAG 3313
Qy      516 GluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHis 535
Db      3314 GAACGGAGCAGTTCCAGCGCACACATATAGCCAAACAAT----- 3352
Qy      536 LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyr 555
Db      3353 -----CCGACAAACCCCTGCCCTCAGACGTGGC----- 3392
Qy      556 GlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeu 575
Db      3383 AGAAAGTCTGTAAACCTCGGAAATGAAG-----CAAGAAATA 3421
Qy      576 AsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595
Db      3422 AAGAGATCATGTCCCAACTCCGGTGGAG-----CTACACAAGGTGACCTTATAC 3472
Qy      596 AsnGlnSerGlyAlaAspPheTyrGly-----AlaGluGlyGluIleTyrPheLys 612
Db      3473 AAGCACTCTGGATGGAGACTTCGGGTTCAGTGTGGCAGATGCGCTGCTG----- 3523
Qy      613 ProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsn 632
Db      3524 -----GAGAAAGCGCTGTATGTC-----AAAAAT 3547
Qy      633 Leu---ProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProphe 648
Db      3548 ATCCGNCAGCTGGGCAGGT-----GATTTGGGGCTTGAAGCCCTAC 3592

```

RESULT 28

BU414889/c

LOCUS

DEFINITION 603550605F1 CSEORBL06 Gallus gallus cdna clone CHEST52012 5', mRNA

ACCESSION BU414889

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

COMMENT

FEATURES

source

BU414889.1 GI:25907560

Gallus gallus (chicken)

Gallus gallus

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 837)

Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,

Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

22335534

12445392

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1..837

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="layer and broiler"

/db_xref="taxon:9031"

/clones="CHEST52012"

/sex="Male and female"

/tissue type="Abdominal fat pad"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="CSEORBL06"

/note="Vector: pBluescript II KS(+); Site 1: EcoRI;

Site 2: NotI; Modification of pBluescript II KS(+)

[Stratagene] vector to accommodate cDNA produced with the

T-trimmed protocol (Construction of uni-directionally

cloned cDNA libraries from messenger RNA for improved 3'

end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387

,624). Cut pBluescript II KS(+) with NotI and EcoRI.

Ligate in double stranded adaptor containing BglI and

BamHI sites [5'ggccggcgagccgcggatccgaaaaaag]

[5'aattcttttttcggatccggggcgacgc]

164 a 230 c 287 g 156 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.84 Length: 837
Score: 107.50 Matches: 60
Percent Similarity: 35.06% Conservative: 35
Best Local Similarity: 22.14% Mismatches: 91
Query Match: 2.66% Indels: 85
DB: 13 Gaps: 13

US-09-936-377-2 (1-758) x BU414889 (1-837)

Qy 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAla----- 22
Db 690 ATCTCGTGAATCTCCTCAATACCTCTTCATGCGCGCTTCTCTAGGAATGTGAGAA 631
Qy 23 ---GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys 41
Db 630 GGGCATGTTTACAGACACCGGGATGTGACAGATTCAGGCCCTTGGTCCAGCTCAGTGCC 571
Qy 42 SerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle 61
Db 570 CTGNAACCGCGCCGCGGGGGATCCACCACTCGGAATCTTCTTGACCCCACTTCTC 511
Qy 62 SerGly-----AspThrLeuArgGlnLysAlaValAsnLeuGly-----AspAlaLeu 77
Db 510 CATCATCTCTCGAAGGTCTTTAGCAGGGTCTTTGTCCAGCTTGGCCTTAAAGACCGCTG 451

Qy	289	GlnLeuTyrProHis-----	LeuLeuThrGluGluAsp	299
Db	257	CAAGTCTACGAACACTGCCCGGGCACATGCTTGCCAGCTCCTGTCTCACTGAAGAAGT	198	
Qy	300	ValAsp-----	301	
Db	197	GTTGAAGGAGTCATCTCTCCCAATGGTCTTGTCACCTGGCATCTGGCCATCAGGCTG	138	
Qy	302	-----TyrAspAsnProGlyLeuSerCysGlyPheHisAsp-----	313	
Db	137	GATGCCATGTTCCAGGCAGTAGAGCTCCAGCAGGCATTGGCGATCTGGACACCCAGCCTG	78	
Qy	314	-----AspAspAlaHisAlaHisAsnGlyLysProTrpIleAspLeu	330	
Db	77	GCCAACGTGGATGGAGATGCACTACGCGATGATACCAACGG-----TTAGATCCC	27	
Qy	331	ArgAsnLysArg	334	
Db	26	AGGGAATCCCG	15	
RESULT 30				
BE123847/c				
LOCUS				
DEFINITION BOD10 BOD-24 Bos taurus cDNA, mRNA sequence.				
ACCESSION BE123847				
VERSION BE123847.1 GI:8517161				
KEYWORDS EST.				
SOURCE Bos taurus (cow)				
ORGANISM Bos taurus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.				
1 (bases 1 to 262)				
Guo,Z. and Zhang,Y.				
Differential expressing genes from bovine cumulus oocyte complexes before and after cultured in vitro				
Unpublished				
CONTACT: Zekun Guo				
LABORATORY OF Developmental Biology				
Northwestern Sci-Tech University of Agriculture and Forestry				
22 Xinqong Road, Yangling, Shaanxi province, 712100, P.R.China				
Tel: 86-029-7098576				
Fax: 86-029-7092176				
Email: devbio@nau.edu.cn; guozekun@163.net.				
FEATURES				
source				
1..262				
/organism="Bos taurus"				
/mol_type="mRNA"				
/db_xref="taxon:9913"				
/sex="female"				
/tissue_type="ovarium"				
/cell_type="cumulus oocyte complex"				
/dev_stages="culture in vitro for 24 hour"				
/clone_lib="BOD-24"				
/note="Sequences obtained by DDRT-PCR method"				
BASE COUNT 88 a 49 c 41 g 84 t				
ORIGIN				
Alignment Scores:				
Pred. No.: 0.387 Length: 262				
Score: 107.00 Matches: 26				
Percent Similarity: 52.22% Conservative: 21				
Best Local Similarity: 28.89% Mismatches: 35				
Query Match: 2.65% Indels: 8				
DB: 10 Gaps: 2				
US-09-936-377-2 (1-758) x BE123847 (1-262)				
Qy	526	ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnIleGlu	545	
Db	259	ACCATCGCTATGAAGTGGGAATCCCAATTTGAAACAGACAAATAATTTCAACCGAT	200	
Qy	546	LeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArg	565	

```

Db 315 -----TTCCGGCGCAACCGCGACGGCTCGATCATCGCGGACGGCATGCGCTCG 268
Qy 124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db 267 GTGCAGGCGCCGAATTTCCACGCCACCCCGACCAACCGATCGTGGAGGTGCTGAAGGGGCCACA 208
Qy 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db 207 TCCCTGCTCTATGGCATCCAGGACCCGGGGGGGGTCAACGTGGTG----- 160
Qy 164 ProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly 183
Db 159 ---AGCAAGAGCCGCACTGCAACAGCCCAACGCCCTGACCCCTCGTGGCTCGGCTCG 103
Qy 184 AsnLeuGluLysLeuThrSerGlyGly 192
Db 102 GCCACAGGGCGCAACGGCAGCGGGCGGC 76

```

```

RESULT 32
BZ573965
LOCUS
DEFINITION msh2_3446.x1 msh Pseudomonas aeruginosa genomic clone msh2_3446,
genomic survey sequence.
ACCESSION BZ573965
VERSION BZ573965.1 GI:27209026
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 874)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source
1..874
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_3446"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT 175 a 278 c 277 g 143 t 1 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.2 Length: 874
Score: 107.00 Matches: 67
Percent Similarity: 35.36% Conservative: 25
Best Local Similarity: 25.77% Mismatches: 101
Query Match: 2.65% Indels: 67
DB: 29 Gaps: 14

```

```

US-09-936-377-2 (1-758) x BZ573965 (1-874)
Qy 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArg----- 45
Db 97 TCCGCGCGGCTGTGTGCGCTGCTCGGACCGCCAGAAACGTGAACGCTCGCTGGAGGAC 156
Qy 46 ---AlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAspLysIle 60

```

```

Db 157 ACAGCAGTCTCTCGGTCTCGCTCACAGCGCACGCGACATTGACCCGAAGCAGAC----- 210
Qy 61 lIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 211 -----CGGCAACGCTCGGTGTCGCGCGGATGATCAATGGCAGTCCCAA 252
Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSer-----AlaProValIleArg 98
Db 253 CGTGGTCTACAC-----CGACTCGTGGCGCGCGGATCATCCGCGCGCAGGACACCCA 306
Qy 99 GlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAsp----- 116
Db 307 GGGCCCGCAACACGGCGCAGAACTGTTCTGGGGCGGCACAGTGCCTGGCGCGCAGCATCAA 366
Qy 117 -----MetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
Db 367 CTGATGGCCACTTACCTGAACACAGATATGTTCTTCGGCGGCGACCTCGGTCTG 423
Qy 133 Ser---GlnGln-ValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAs 151
Db 424 GGAGCTCGACAGCATCGAGGTGTTCCGCGCGCGCGCAGACCACTCCAGGGCGCCAAACGC 483
Qy 151 nValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGl 171
Db 484 CATGCCCGGGCGGATC---ATCGTCAACACCAAGGACCGACCTTCAGCCCCGAGGCTGG 540
Qy 171 yValSerGlyGluLeuGlyLeuArgLeuSer-----SerGlyAs 184
Db 541 CTACCAAGGGGGGATGTCAGCTACCATTCGCGCGCAGATGTATCGGAAGTCCGGCCC 600
Qy 184 nLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHi 204
Db 601 GTTGGCAAGACTTTGGCGGGCGCC-----GGCGGGGGG----- 634
Qy 204 sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLy 224
Db 635 -ACTACGCGGTGCGCACACCTTCATTGATACGCAACCCCGAAGTTCACGACAGGGG 693
Qy 224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
Db 694 GCACCGACCGAGGACTTCG-----GGCCCTTAAGCGCGCGGCCAAGCTG 738
Qy 244 pArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro 263
Db 739 G-----TTGGCTTGCCA 751

```

```

RESULT 33
BZ065834/c
LOCUS
DEFINITION Fgr 8_K04_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
sequence.
ACCESSION BZ065834
VERSION BZ065834.1 GI:22506123
KEYWORDS EST.
SOURCE Gibberella zeae
ORGANISM Gibberella zeae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 634)
AUTHORS Trail,P., Xu,J.-R., San Miguel,P., Halgren,R.G. and Kistler,H.C.
TITLE Analysis of expressed sequence tags from Gibberella zeae (anamorph
Fusarium graminearum)
JOURNAL Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE 22508120
PUBMED 12620255
COMMENT Contact: Frances Trail
Department of Plant Biology
Michigan State University
East Lansing, MI 48824, USA
Tel: 517 432 2939
Fax: 517 353 1926
Email: trail@msu.edu

```

```

Plate: 8 row: K column: 04.
FEATURES source Location/Qualifiers
1. .634
/organism="Gibberella zeae"
/mol_type="mRNA"
/db_xref="NRRL 31084"
/clone_lib="Nitrogen-starved mycelia"
/notes="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 128 a 231 c 154 g 121 t
ORIGIN
Alignment Scores:
Pred. No.: 1.56 Length: 634
Score: 106.50 Matches: 67
Percent Similarity: 38.70% Conservative: 22
Best Local Similarity: 29.13% Mismatches: 78
Query Match: 2.64% Indels: 63
Gaps: 13
US-09-936-377-2 (1-758) x BU05834 (1-634)
Qy 18 ThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThr 37
Db 587 ACTCTTTACAGACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 528
Qy 38 ValVal-----GlyLysSerArgProArgAla 46
Db 527 GCCAGCAGCACCCTTGGCTCTCGGAGCCTTGGTAGGGTGGCGGGGGTGGTCT 468
Qy 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu 66
Db 467 GGTGGCGGCTCTCTCAGCAGCGGTGGTGGGAATGATGGAGCGAGCGAGTGGTGGCAGC 408
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db 407 GCGGTGGTCTCAACACCTCGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 348
Qy 87 GlnTyrglyGlyAlaSerAlaProValIleargGlyGlnThrGlyArgArgIle---- 105
Db 347 -----GGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 294
Qy 106 -----LysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 293 GGGGGGTTCTGCGAGGTTCTC-----GGTAGCGGGAAGGACCTCGTTGATGGC 243
Qy 121 SerProAspHisAlaIleMetValAspThr----- 130
Db 242 AACGTGGTACCACTCGTCAATGACACAGAGGTAGCTTGGATCGGACAGCGCTCAA 183
Qy 131 AlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrrSerSerGly 150
Db 182 GCCTTCGACACACAGCGAGGTC-----GGTCTGTGTCAC----- 147
Qy 151 AsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
Db 146 -----CTGGTCTC-----GCTGGGATGGCTTGTCTGAGGCAAG-----AAT 108
Qy 171 GlyValSerGlyGluLeuGlyLeuArgLeuSer---SerGlyAsnLeuLeuGluLysLeuThr 189
Db 107 GCGCACTTGGGACGTCGCGACAGAGACTGAGCCTGAGCGGCAAC----- 63
Qy 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrr 209
Db 62 AGCGCGAGACAAACGAAAGCGAGG-----GTACTTCATTTTGAATGTGTGTTT 12
Qy 210 ArgLysSerGlyAspTyrAlaValProArg 219
Db 11 -----GTTCTCTGT 3
CA324304 880 bp mRNA linear EST 26-NOV-2002

```

DEFINITION UI-M-FY0-ccp-e-01-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
IMAGE: 6822602 5', mRNA sequence.
CA324304
VERSION CA324304.1 GI:24542402
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 880)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. .880
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6822602"
/tissue_type="whole brain"
/dev_stage="embryo 13.5, 14.5, 16.5, 17.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoRI; Site_2: NotI; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a NotI site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoRI adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the NotI site and the polyA tail is AGCAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 242 a 230 c 235 g 168 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 2.51 Length: 880
Score: 106.50 Matches: 57
Percent Similarity: 32.96% Conservative: 32
Best Local Similarity: 21.11% Mismatches: 108
Query Match: 2.64% Indels: 73
Gaps: 14

US-09-936-377-2 (1-758) x CA324304 (1-880)

Qy 102 GlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 145 GGGCGCCAGTCACGACCCAGCCAGCAGTCACCCGCCCTCTGAGTCTCAGTCACAGTCA 204
Qy 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIle---LeuArg 140
Db 205 CCAGGCCGAGAGGTGATGACTATGCTCAGCCCGAGCGCATGGTAGAGTGAGCCTTCAT 264
Qy 141 GlyProValThrLeuLeuTyrrSerSerGlyAsnValAlaGlyLeuValAspValAla--- 159

RESULT 34
CA324304
LOCUS

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'-GAGAGAGATCTTCGATTAATTAATTAATCCCCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified pluescript KS(+) after bulk excision from Lambda F1C I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

FEATURES

```

RES
source
Location/Qualifiers
1. .2531
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4931400F03"
/db_xref="MGI:1907112"
/db_xref="taxon:10090"
/clone="4931400F03"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"

```

misc feature

```
misc_feature
1. .2531
/note=similar to GLUTAMATE RECEPTOR INTERACTING PROTEIN
1A-1 [Mus musculus] (SPTR|0925t6, evidence: PASY,
76.2%ID, 98.4%length, match=3789)"
/db xref=MGI:1921303"
```

BASE COUNT	749 a	627 c	619 q	536 t
BASE COUNT	749 a	627 c	619 q	536 t

ORIGIN

Alignment Scores:

Pred. No.:	13	Length:	2531
Score:	106.00	Matches:	151
Percent Similarity:	35.2%	Conservative:	86
Best Local Similarity:	22.4%	Mismatches:	227
Query Match:	2.63%	Indels:	211
DB:	11	Gaps:	38

US-09-936-377-2 (1-758) x AK016420 (1-2531)

30	SerValGlyLeuGluThrValThrValValGlyLysSerArgProAraGalaThrSerGly	49
322	TCTTTAGCCTCCAGCACTGTGGGGTGGCTGGCCAG-----	357
50	LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArg-----	67
358	GTCTTCACACTGAACC-----ACAGAGTTGTCTGACGGCTGACCTGTCAACGGGCTTC	414
68	-----GlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIle	83
415	GGAAATCCAACTGCAGGGCAGCGTGTTCACAGACAGCGCTCTCTCTCCGCGCTCTGATT	474
84	HisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGly-----	100
475	TCCTATATTGAA-----GCTGACAGCCCGACAGAGATGCTGTGTCTACAG	522
101	ThrGlyArgArgIleLysValLeuAsnHisGlyGlyLeuThrGlyAspMetAlaAspPhe	120
523	ATTGGAGCAGAGTCTATGCCATTAAAT-----GGAAATCCCAACAGACAGACCGCTTC	576
121	SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg	140
577	GAGGAAGCCCAATCAACTCTCTGAGAGACTCTTCCATCAGCAAAA-----	621
141	GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp	160
622	-----GTCACACTGAATATCGAG-----TTTGTATGTTGCAGAG	654
161	GlyLysIleProGlu-----LysMetProGluAsnGlyValSerGly	174
655	TCTGTCAATCCCAAGTAGTGGAACTTTTCATGTAAACTGCCCAAGACAC---ACGGTG	711
175	GluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsn	194

712 GAACTTGGGAATAACCATCAGTTCCCATCCAGTAGAATAACCGGGAGACCCCTTGTCAATT 771

195 ILeGlyLeuGlyLysAsnPheValLeuHisThrGluGly-LeuTyrArg----- 210

772 TCAGATATCAAGAAAGGCAGTGTGGCTCACAGAACCGGAACCTCTGGAACCTGGGAGACAAA 831

211 -----LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysAr 225

832 CTGCTTGCATAGATAACATCCGCTCGATAACTGTTCCTCA-----TGAAGAGT 879

225 gLeu-ProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpA 245

880 GCGGTCCAGATCTCCACAGCATGTGAAGACCTGGTGAA-----GCTCAAAATCCGCAAA 933

245 rGlysArgPheTyrArgArgThrTyrSerAspArgArgAspGln----- 259

934 GATGAAGATAACTCAGACGACGACAGAGATTCGGAGCGATTATTACACGGTGGAGCTG 993

260 -----TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleI 277

994 AAGCGCTATGGGGGGCCCT----- 1013

277 leuTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrG 297

1014 --TGCGATACAAATTCGTGAAGCTGAAGCGGTTTGATCTATTATCATCTC----- 1064

297 luGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspA 317

1065 -----GAGC 1068

317 laHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLys----- 333

1069 CTCACATAAGGGGATTAGCTGAAGACTGGAGCGATCCACATCGGAGATGAGATCCTA 1128

334 -----ArgTyrGluLeuArgAla--GluTrpLysGlnProPhe--ProG 347

1129 GCCATCAATAGCAGCAGCTTGAAGGGGAAGCCCTCTGAGTGAAGCATCTCCTGCTCCAG 1188

347 lyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaG 367

1189 AT-----GGCAGGAGAGACTCTGCCTCTGAAATAATTAAGAAA 1224

367 lyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgH 387

1225 CAGACAGATGGCTCAACTTT-----TCAGACTTCAGATACATATTAACAC 1269

387 isGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerS 407

1270 ACCT-ATGATTTGGA-----GGAGTCCAA-----AGCAAGA-A 1300

407 erAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValG 427

1301 CCAGCTGTCTC-----CCAGTCCCAAGCCT-----CGAAGCC 1333

427 lnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyG 447

1334 AGACGTACCCAGATGTGGCTGAGTAATGAAGACTGGGATCGATCCACAGCCAGTGCTCT 1393

447 lyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArg- 466

1394 TTGTAGGGGGCTTCT-----GACAGTGCAGATGCTGAACAAG 1429

467 --GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla---- 484

1430 AGGAAACTTCTGGTCTCAAGCATTTGAGACCTGGAGACCTCGCGCCAGTCCGGGGATCC 1489

485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysL 501

1490 TGACAGAGCTTGAGGCACAATCATGTCTGGGGAGT---ACTATGAGTTTGAATCATGAGT 1546

501 euSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisG 521

1547 CTCCAATGGCTCCAGTCCAGCTGGGCGCAGCAGCCAGCTTCAGGAACCGGACGAGTTCAC 1606


```

QY      521  lylyshisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgS 541
DB      1607  GGCCACATATAGCAACAACT-----CGCA 1633
QY      541  erAsnAsnIleGluLeuAlaLeuGlyTyrgluGlyAspArgTrpGlnTyrAsnLeuAlaL 561
DB      1634  GCAACACCCCTGCCTCAGAGCTGGC-----AGAAAGTCGTAAACC 1675
QY      561  euTyrgArgAsnArgPheGlyAsnTyrlfTyraGlnThrLeuAsnAspGlyArgGlyP 581
DB      1676  TGGGAAATGAAG-----CAAGAAATAAAGAGATCATGTCCC 1714
QY      581  rlyysSerIleGluAspAspSerGluMetIlyslLeuValArgTyrAsnGlnSerGlyAlaA 601
DB      1715  CAACCTCGGGAG-----CTACAAAGGTGACCTTATACAGGACTCTGGCATGG 1765
QY      601  spPheTyrgly-----AlaGluGlyGluIleTyrlPheLysProThrProArgTyra 618
DB      1766  AGGACTTCGGGTTCCAGTGTGCAGATGGCTGCTG----- 1800
QY      618  rgIleGlyValSerGlyAspTyrgValArgGlyArgLeuLysAsnLeu---ProSerLeuP 637
DB      1801  -----GAGAAAGGGCGTGTATGTC-----AAAAATATCCGCCAGCTGGGC 1840
QY      637  roGlyArgGluAspAlaTyrglyAsnArgProphe 648
DB      1841  CAGGT-----GAGTGTGGGGCTTGAAGCCCTAC 1869

RESULT 36
BZ564107
LOCUS   BZ564107.1 686 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_4573.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ564107.1 GI:27189244
VERSION   BZ564107.1
KEYWORDS  Pseudomonas aeruginosa
SOURCE    Pseudomonas aeruginosa
ORGANISM  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
REFERENCE 1 (bases 1 to 686)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
          Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol., (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.

FEATURES             Location/Qualifiers
     source           1..686
     organism="Pseudomonas aeruginosa"
     /mol_type="genomic DNA"
     /strain="2-164"
     /db_xref="taxon:287"
     /clone="pacs2-164_4573"
     /clone_lib="pacs2-164"
     /note="clinical isolate 2-164 Whole genomic shotgun
     library."

BASE COUNT  127 a  227 c  229 g  103 t
ORIGIN
Alignment Scores:
Pred. No.:      2.22      Length:      686
Score:          105.50    Matches:      31
Percent Similarity: 47.62%  Conservative: 9
Best Local Similarity: 36.90%
Query Match:    2.61%
DB:             29
US-09-936-377-2 (1-758) x BZ563690 (1-1359)

```

```

Query Match: 2.61% Indels: 5
DB:          29      Gaps: 3
US-09-936-377-2 (1-758) x BZ564107 (1-686)
QY      678  AlahenLeuAspTyrgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArg 697
DB      144  GCGGGGTGACGTCGACACACCGCCCTCGCAGCACCGGTGCGGAGGAAGAACTCTCC 203
QY      698  ThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrgArgAsnThrArgTyrgly 717
DB      204  ACCGACGGCTAC-----ACCACCTCGCGCCCTCGGCTACACTTCGACCTCGGC 257
QY      718  GluTrpAsnTrp-----TyrVallyAlaAspAsnLeuLeuAsnGlnSerValTyraLa 735
DB      258  GAGACCCCTGGTGGCTTCTGTCAGGGGACCAACCTGACCAACGACGCTGCGCTAC 317
QY      736  HisSerSerPheLeuSerAsp---ThrProGlnMetGlyArgSerPheThrGlyGlyVal 754
DB      318  GCAGTTTCGATCTGCGCGACCGGTGCGCGACGGGACGCGGCATCGAGGCGGGGTG 377
QY      755  AsnVallyysPhe 758
DB      378  AAGGTGGCGGTTCC 389

RESULT 37
BZ563690
LOCUS   BZ563690.1 1359 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_4378.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ563690.1 GI:27188088
VERSION   BZ563690.1
KEYWORDS  Pseudomonas aeruginosa
SOURCE    Pseudomonas aeruginosa
ORGANISM  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
REFERENCE 1 (bases 1 to 1359)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
          Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol., (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.

FEATURES             Location/Qualifiers
     source           1..1359
     organism="Pseudomonas aeruginosa"
     /mol_type="genomic DNA"
     /strain="2-164"
     /db_xref="taxon:287"
     /clone="pacs2-164_4378"
     /clone_lib="pacs2-164"
     /note="clinical isolate 2-164 Whole genomic shotgun
     library."

BASE COUNT  271 a  385 c  278 g  422 t  3 others
ORIGIN
Alignment Scores:
Pred. No.:      5.96      Length:      1359
Score:          105.50    Matches:      31
Percent Similarity: 47.62%  Conservative: 9
Best Local Similarity: 36.90%
Query Match:    2.61%
DB:             29
US-09-936-377-2 (1-758) x BZ563690 (1-1359)

```



```

DEFINITION 603111672F1 CSEQCHL12 Gallus gallus cdna clone CHEST59018 5', mRNA
sequence.
ACCESSION BU106765
VERSION BU106765.1 GI:25309076
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE 1. (bases 1 to 858)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.R., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
MEDLINE Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED 22335534
COMMENT 12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..858
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hixen"
/db_xref="taxon:9031"
/clone="CHEST59018"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSQCCHL12"
/note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the
7-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BspI and
BamHI sites [5'ggcgcgtgacgccggatccgaaaaag]
[5'aattcttttcggtccggtgacgc]"
BASE COUNT 164 a 242 c 288 g 164 t
ORIGIN
Alignment Scores:
Pred. No.: 3.46 Length: 858
Score: 105.00 Matches: 44
Percent Similarity: 38.89% Conservative: 26
Best Local Similarity: 24.44% Mismatches: 68
Query Match: 2.60% Indels: 42
DB: 13 Gaps: 8

US-09-936-377-2 (1-758) x BU106765 (1-858)
QY 15 LeuileAsnThrProLeuLeuAla-----GlnAlaHis 25
Db 687 CTCCTCAATACCTTCTCATGCGCGTTCTGTTAGATGTTGAGAGGGCATGTTAC 628

QY 26 GluThrGluGlnSerValGlyLeuGluThrValThrValGlyLysSerArgProArg 45
Db 627 AGCACCAGGGAGTGTGACCAAGATTCAGCGCTTGTGTCAGCTCAGTGCCTGAAACCGGCC 568

QY 46 AlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGly----- 63
Db 567 CCACGGCGGGGAATCACCACTGGAACCTTCTTGGACCCCAATTCATCATCATCTCTC 508

QY 64 AspThrLeuArgGlnLysAlaValAsnLeuGly-----AspAlaLeuAspGlyValPro 81
Db 507 GAAGGCTTTAGCAGGGTCTTCTCCAGCTTGGCCCTTAAGACCGCTGCGGGGTGGCT 448

```

```

QY 82 GlyIleHis-----GlyGlyAlaSerAlaProValIleArgGlyGlnThrGly 88
Db 447 GGGCTCCGCCGTACAGGGGTGCCCTCTCTCCACCGAGTCTTGAAGCCACCGTTCAGCAC 388

QY 89 Gly-----GlyGlyAlaSerAlaProValIleArgGlyGlnThrGly 102
Db 387 GGAGACCTCGCGTGCCTCCAAAGGCTCGGAACATCCACACGGCGGGGGGCATAGAAGT 328

QY 103 ArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 327 GCCAGCTCGTCCGCTCATACACAC---CACGTGGTGTGCTGCTGACCCCGGCG 271

QY 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnValGluLeuLeuArgGlyPro 142
Db 270 CCCAC---GTAGTCGGCAAGTGGCCCTCGTCGCACATGAATCATACGGGA--- 217

QY 143 ValThrLeuLeuTy-SerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLys 162
Db 216 -----GGACTTATCCCGGCACTCTCTGATGTCGAAGAAGGACGC 178

RESULT 40
AL969106 932 bp mRNA linear EST 27-NOV-2002
LOCUS AL969106 XGC-gastrula Silurana tropicalis cdna clone TGasi27p16 5',
DEFINITION mRNA sequence.
ACCESSION AL969106
VERSION AL969106.1 GI:25792701
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 932)
AUTHORS Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2002
JOURNAL Unpublished
COMMENT Contact: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TGasi27p16.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
Constructed by Aaron M. Zorn.
FEATURES
Location/Qualifiers
1..932
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGasi27p16"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dT primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT 271 a 185 c 242 g 233 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 4.39 Length: 932
Score: 104.50 Matches: 80
Percent Similarity: 38.18% Conservative: 46
Best Local Similarity: 24.24% Mismatches: 117
Query Match: 2.59% Indels: 88
DB: 9 Gaps: 16

US-09-936-377-2 (1-758) x AL969106 (1-932)

```

```
Qy 233 PheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThr 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16 TTTCCTAGGACCAATGTAGA-----TTCTTTCTTACC 48
Qy 253 TyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCys 272
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 TTGGCAGATTGTATTTCCTGTTGGT-----GACAAATAAT 84
Qy 273 HisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeu----- 288
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 AATGCCATCTTAAGTTTGTAGCAGATGAGCCAGATGCTGGATCCTTATCTCATAAAGGAATG 144
Qy 289 GlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSer 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 GATATATATGGTACTTCTGCGCGTGAAGTGCAGTGCAGAGATGTGGAATAATCTTGGC 204
Qy 309 CysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProIle 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 TGCGCTTGTTCAAATATATCTGATCAGCATGCT-----GAGCCATGGGTG 249
Qy 329 AspleuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPhe 348
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 GTTTCGGTCCACAGCTTTTATACAAAGAGACACTCCAGAGCCCTGTATTAGTGCC 309
Qy 349 GluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAsp 368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 AAGGCT-----ATCCAGCTGAACAGCAATAGTGTCAAGCCTTACTCTTAAAGGGGCA 363
Qy 369 AlaValGluAsnPhePheAsnGlnThrGlnAsnAlaArgIleGluLeuArgHis--- 387
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 GCATTAAAGAACATG-----GGAAGGTCAGGAAGCAATAATTCAATTCGAGAGCA 417
Qy 388 -----GlnProIleGlyArgLeuLysGlySerTrpGlyVal-----GlnTyrLeu 402
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 ATTCGATTAGCACCTTGT---CGCTTGACTGCTGTGAAGTCTCATTGAATGCTATTTG 474
Qy 403 GlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeu 422
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 -----GCATCCAGCAGTGTGCGGAGCCATGGTGATG 507
Qy 423 AspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPhe 442
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 GCCAACAAATGTC-----TACAAACCCCTTGA-----GCCACGCACAAACTCTA 552
Qy 443 ThrLeuGluGlyGlyValArgValGlu-----LysGlnLysAlaSerIleArg 458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
553 ACTCTGTGGCACTGTGCTGGAGATCCTGTACCAGAGAGAGCAAAACATTG 612
Qy 459 TyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGly 478
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
613 TTAGACAAAGCGCTCAITTCAGAGACCGGATTATATAAAGGCTGTGTGAA--GAAAGCGGA 671
Qy 479 AlaHisArg----- 481
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
672 GTTACTAAGTCGAGAGAGAGATGATGAGAGGAATTGCCCTGCTGCGTATACCTTGC 731
Qy 482 -----GlnThrAlaArgSerPheAla----- 488
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
732 AAACCAAGTGACTGCGTACTACTACATCGATGCTAGGAGATTTCTGGTGGCAGTTAATGA 791
Qy 489 LeuSerGlyAsnTrpTyrPheThrProGlnHisLys-LeuSerLeuThrAlaSerHisG1 508
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
792 GTTTCAGGAAGC---TATGGATCAGTACAGATCGCCTCAGTTTGGATCCAAATGACCA 848
Qy 508 nGluArgLeuProSerThrGlnGluLeu 517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
849 AAAATCCTTGGAGGAATGCAGAAATG 876
```

Search completed: November 17, 2003, 19:28:03
Job time : 2707 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:00:13 ; Search time 36 Seconds
(without alignments)
3843.891 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTLKPIVLSILLINTPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 18259486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3845	95.3	764	15	US-10-066-551-2
2	216	5.4	713	11	US-09-801-451A-4
3	214.5	5.3	720	15	US-10-066-551-3
4	202	5.0	713	11	US-09-801-451A-2
5	198.5	4.9	687	12	US-10-181-319-2
6	185	4.6	660	12	US-10-238-075-732
7	181	4.5	654	12	US-10-238-075-527
8	180.5	4.5	914	15	US-10-043-344-11
9	177.5	4.4	713	12	US-10-238-075-263
10	163	4.0	913	15	US-10-043-344-5
11	161.5	4.0	912	15	US-10-043-344-7
12	161.5	4.0	912	15	US-10-043-344-9
13	160	4.0	540	12	US-10-238-075-721
14	156	3.9	971	15	US-10-098-808-2
15	155.5	3.9	682	12	US-10-238-075-1077

16	155	3.8	908	15	US-10-043-344-94	Sequence 94, Appl
17	154	3.8	813	9	US-09-815-242-5085	Sequence 5085, Ap
18	154	3.8	911	15	US-10-043-344-107	Sequence 107, App
19	151.5	3.8	915	12	US-10-240-218-6	Sequence 6, Appli
20	150.5	3.7	915	9	US-09-332-226-2	Sequence 2, Appli
21	150.5	3.7	915	15	US-10-043-344-96	Sequence 96, Appl
22	147.5	3.7	909	9	US-09-332-226-4	Sequence 4, Appli
23	147.5	3.6	911	15	US-10-043-344-95	Sequence 95, Appl
24	141.5	3.5	729	9	US-09-815-242-10132	Sequence 10132, A
25	135	3.3	791	10	US-09-881-752A-334	Sequence 334, App
26	127	3.1	815	9	US-09-815-242-5106	Sequence 5106, Ap
27	123	3.1	793	12	US-09-881-752A-362	Sequence 362, App
28	122.5	3.0	767	10	US-09-882-227-78	Sequence 78, Appl
29	122.5	3.0	1590	15	US-10-180-326-1	Sequence 1, Appli
30	120.5	3.0	753	12	US-10-238-075-1277	Sequence 1277, Ap
31	118.5	2.9	1475	9	US-09-740-274-2	Sequence 2, Appli
32	117.5	2.9	422	9	US-09-815-242-5247	Sequence 5247, Ap
33	117	2.9	978	9	US-09-815-242-5456	Sequence 5456, Ap
34	117	2.9	1001	9	US-09-815-242-12686	Sequence 12686, A
35	115.5	2.9	981	9	US-09-815-242-12211	Sequence 12211, A
36	112	2.8	631	15	US-10-260-877-62	Sequence 62, Appl
37	111	2.8	1197	10	US-09-738-626-6493	Sequence 6493, Ap
38	109.5	2.7	818	14	US-10-055-364-44	Sequence 44, Appl
39	108	2.7	121	14	US-10-022-461-5	Sequence 5, Appli
40	108	2.7	907	12	US-10-130-973A-5	Sequence 7943, Ap
41	107.5	2.7	430	15	US-10-156-761-7943	Sequence 30, Appl
42	106	2.6	511	15	US-10-122-706-30	Sequence 5, Appli
43	106	2.6	970	12	US-10-331-061-5	Sequence 30, Appl
44	106	2.6	1052	12	US-10-130-973A-17	Sequence 17, Appl
45	106	2.6	1172	15	US-10-122-706-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-066-551-2
; Sequence 2, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-2

Query Match	95.3%	Score	3845	DB	15	Length	764
Best Local Similarity	95.9%	Pred. No.	0				
Matches	730	Conservative	7	Mismatches	18	Indels	6
Oy	1	MAQTLKPIVLSILLINTPLLAQAHETEQSVGLTETVVGKSRPRATSGLLHTSTASDKI	60				
Db	7	MAQTLKPIVLSILLINTPLLSQAHQEQSVGLTETVVGKSRPRATSGLLHTSTASDKI	66				
Oy	61	ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGTGRRIKVNLHHGTGMADF	120				

Db	67	ISGDTLRKAVNLDGVALDGVPGIHA	SGVGGASAPVIRGQTGRRIKVLNHHG	ETGDMADF	126		
QY	121	SPDHAIWDTAL	SOQVEILRG	SPVTLTLLSSGNVAGLVDVADGKI	PEKMPENGVS	GELGURL	180
Db	127	SPDHAIMVDSAL	SOQVEILRG	SPVTLTLLSSGNVAGLVDVADGKI	PEKMPENGVS	GELGURL	186
QY	181	SSGNLEKLTSGG	INTGLGKNFVLHTEGLYKSGDYAVPRV	RNILKRLPDS	PRPFANGQRA	240	
Db	187	SSGNLEKLTSGG	INTGLGKNFVLHTEGLYKSGDYAVPRV	RNILKRLPDS	---	HAUSQTS	243
QY	241	V-LGW--RKR	FVRRYTSRRDQYGLPAHSHEYDCHADI	TWQSLINKRYLOLYPHLLTE	297		
Db	244	IGLSWVGKGF	TAAYSRRDQYGLPAHSHEYDCHADI	TWQSLINKRYLOLYPHLLTE	303		
QY	298	EDVDYDNPGLS	CGFHDDDAHAHANGKPM	IDLANKRYELRAEWKQPF	PGFEALR	VHLNR	357
Db	304	EDIDYDNPGLS	CGFHDDDAHAHANGKPM	IDLANKRYELRAEWKQPF	PGFEALR	VHLNR	363
QY	358	NDYHDEKAGDA	VENFNNQTONARIELRHOP	IGRLKSGWGYLVQOKSSALSATSEAVK	417		
Db	364	NDYRDEKAGDA	VENFNNQTONARIELRHOP	IGRLKSGWGYLVQOKSSALSATSEAVK	423		
QY	418	QPMLLDNKVQHY	SPFGVQANWNP	TLBGGVVRVKQKASTRYD	KALIDRENYKQ	PLPDL	477
Db	424	QPMLLDNKVQHY	SPFGVQANWNP	TLBGGVVRVKQKASTRYD	KALIDRENYKQ	PLPDL	483
QY	478	GAHQQTARSA	FALSGNWWYFTPOHKLSLTASHQERL	PSTQELYAHGKHVATNTFV	GVGNKHLN	537	
Db	484	GAHQQTARSA	FALSGNWWYFTPOHKLSLTASHQERL	PSTQELYAHGKHVATNTFV	GVGNKHLN	543	
QY	538	KERSNNIELALGYEGDRQWYNAL	YRNRFGNYIAQTLNDGRGPKS	IEDDSEMKLYR	VYRNO	597	
Db	544	KERSNNIELALGYEGDRQWYNAL	YRNRFGNYIAQTLNDGRGPKS	IEDDSEMKLYR	VYRNO	603	
QY	598	SGADFYGABGEI	YFKPTPRYIGVSGDYVRGLKNL	PSLPGREDAYGNRPFIA	QDQONAP	657	
Db	604	SGADFYGABGEI	YFKPTPRYIGVSGDYVRGLKNL	PSLPGREDAYGNRPLIA	QADQONAP	663	
QY	658	RVPAARLGHV	KASLTDRIDANLDYRYFAONKLAR	YETRTPGHMLNLGANYR	NRTRYG	717	
Db	664	RVPAARLGHV	KASLTDRIDANLDYRYFAONKLAR	YETRTPGHMLNLGANYR	NRTRYG	723	
QY	718	ENRWYVKADN	LLNQSVYAHSSFLSDTPQ	MGRSFTGGVNVKF	758		
Db	724	ENRWYVKADN	LLNQSVYAHSSFLSDTPQ	MGRSFTGGVNVKF	764		

RESULT 2

```

US-09/801-451A-4
Sequence 4, Application US/09801451A
Publication No. US20030096368A1
GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
Beucher, Margaret
TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule
and Vaccine
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Inclone Systems Incorporated
STREET: 180 Varlick Street
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/801,451A
FILING DATE: 08-Mar-2001
CLASSIFICATION: <Unknown>

```

Db 617 NPNLEIGHGRVY-----QKATGSILAAGQKDRDGLNENVRVQGFVY 659
 QY 674 DRIDANLDYVYRPAQNKLYETRTGCHMLNLCANVRRNTRYGEWNVWVADNLLNSQV 733
 Db 660 -----DVPANWKLKGLKOTLN-----VNLNVN-----NVDFKPY 687
 QY 734 YAHSSFLSDT-POMGRSFTGGVNVKF 758
 Db 688 YPHSQRTWNTLPGVGRDVRGLGVNYKF 713

RESULT 3
 US-10-066-551-3
 ; Sequence 3, Application US/10066551
 ; Publication No. US20030100071A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Apicella, M. A.
 ; APPLICANT: Edwards, J. L.
 ; APPLICANT: Gibson, B. W.
 ; APPLICANT: Scheffler, K.
 ; APPLICANT: Brown, E.
 ; TITLE OF INVENTION: Vaccine and compositions for the
 ; TITLE OF INVENTION: prevention and treatment of Neisserial infections
 ; FILE REFERENCE: 875.045051
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: US 60/344,452
 ; PRIOR FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: US 60/310,356
 ; PRIOR FILING DATE: 2001-08-06
 ; PRIOR APPLICATION NUMBER: US 60/266,070
 ; PRIOR FILING DATE: 2001-01-31
 ; NUMBER OF SEQ ID NOS: 12
 ; SEQ ID NO 3
 ; LENGTH: 720
 ; TYPE: PRT
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-066-551-3

Query Match 5.3%; Score 214.5; DB 15; Length 720;
 Best Local Similarity 21.6%; Pred. No. 3.6e-11;
 Matches 186; Conservative 101; Mismatches 315; Indels 259; Gaps 45;

QY 16 INTPL-----LQAQTEQSGVLETVTVVVKSRPRATSGLLHTSTASDKIIS 62
 Db 1 MNTFLFLLSLSLAAGFAAENNAKVLDITVIVKGRQ-----GSKRTIV- 50
 QY 63 GDTLRQK-----AVNLGDALDGVPGIHASQYGGGASAPVI-----RGQTGRRIKY 107
 Db 51 --TLQXDESTATDMRELLKEPSI---DFGGNGTSGQLTLRGWGQNSVDIKVDNAYS 105
 QY 108 ---LNHGETGDMADFPDHAIMVDYDALSQOVELTIRGPVILLYSSGNVAGLVADVADKIP 164
 Db 106 SQILYHQGR-----FIVDPALVKKVVSQKQ-----AGSASAGIATNGAII 146
 QY 165 EKMPF-----NGVSGELGLRSSG--NLEKLTSGGINIGLKNFVLHTEGLY-----RKSG 213
 Db 147 AKTVDAQDLLKGLDKNMGVRLNSGFASNEGVSYGASVFGKEGNF-----DGLFSNNDK 202
 QY 214 DYAVPR-VRNUL---KRLPDSP-----RRFANGQHRVAGVGRKRYFR----- 251
 Db 203 DYEAGKGFNVNGKTVFYSALDKRSYLAKIGITTFGDDHRRVILSHMKDQHRGIRTVREE 262
 QY 252 -TYSRRDQYGLPAHSHEYDDCHADIIWOKSLINKRVLYLPHLLTBEVDYDNPGLSCG 310
 Db 263 FTVGKSSRINDROAPAYR-----TTQSTNLTATGKVLG 299
 QY 311 FHDDDDAHAAHAKPFDLANKRY---ELRAEWKQPPGPEALRVHLNRNDYHHDKAG 367
 Db 300 FVEKLDANAY-----LEKERYSADDSGTGVAGNVKGNPHHTTRITRGANFNFDGRLA 351
 QY 368 DAVENFFNQOT-QNARIELRQ---PIGLKSGMGVQVYLGOKSSALTSATSEAVKQPM--- 420

Db 352 E-----OTLLKYGINVRHQEIKPOAFLNSKFSIPTBEKN-----GQKVDKPMEQ 397
 QY 421 -----LLDNKVQHYSPFGEQA-NMDNFTLEGSVRVEKOKASIRYDKALIDR 466
 Db 398 MKDRADEDTVHAYKLSPTTDTGTVVEAHDIGDFTLTGGLRYDRFKVK-THDGKTVSS 456
 QY 467 ENYTKQPLDLAGHQRARSFALSNGWYFTPOHKLSTASHQERLPSTOELYAHGKH--- 523
 Db 457 SNLN-----PSPGVWQPH-----HWSFSASHNY---ASRSPRL---YDALQTHCKRGII 501
 QY 524 -VATNTFEVGNKHLNKNRSNNIELALGVEGDRWQYNLALYNNRGNVYIYAQLNDGRPK 582
 Db 502 SIADGT-----KAERARNTGTFNNDGTFAAN-----GSYFW-OTIKDALANP 544
 QY 593 STEDDSEMKLVYNSQADFYGAE-GEIYFKPTPRYRIGVSGDYVRGLKNLPLSPGRED 641
 Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASVYRTGGLTAKGVVS-----HSKPRFYD 593
 QY 642 AYGNRPFTIAODDQNAFVPAARLGFHLKASITDRI-DANLDY-YRVFAONKLARYETRT 699
 Db 594 THKDKLLSANPEF-----GAQVGRTWASLAYRFQNPENLEIGWR-----GRYVOKAT 640
 QY 700 GHMMLNLCANVRR-----NTRYGEW-----NWYVKADNLLNQSVYAHSS 738
 Db 641 G-SILAAQCKDKRKNLENVVRKGFVNDVPANWPLGKDTLNVLSVNNVFNKPYPHSQ 699
 QY 739 FLSDT-POMGRSFTGGVNVKF 758
 Db 700 RWTNLTGVRDVRGLGVNYKF 720

RESULT 4
 US-09-801-451A-2
 ; Sequence 2, Application US/09801451A
 ; Publication No. US20030096368A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sparling, P. Frederick
 ; TITLE OF INVENTION: Isolated PrpB Nucleic Acid Molecule
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ImClone Systems Incorporated
 ; STREET: 180 Varick Street
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10014
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION NUMBER: US/09/801,451A
 ; FILING DATE: 08-Mar-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/418,964
 ; FILING DATE: 07-APR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gallagher, Thomas C.
 ; REGISTRATION NUMBER: 37,066
 ; REFERENCE/DOCKET NUMBER: SPA-3-P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 645-1405
 ; TELEFAX: (212) 645-2054
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 713 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-801-451A-2

```

Query Match          5.0%; Score 202; DB 11; Length 713;
Best Local Similarity 21.6%; Pred. No. 5.6e-10;
Matches 187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;

QY 11 LSILLINTPL-LQAQHEQSVGLTETVVVVGKSRPRATSGLLTSTASDKISGDTLRQK 69
Db 8 LSLLSLTLAAGFAHAENNAVALDVTYVKGDRQ-----GSKRTNIV---TLQOK 55

QY 70 ----AVNLGALDGVPGIHASQVGGASAPVI-----RGOTGRRIKV-----LNHH 111
Db 56 DESTATDMRELLKEPSI---DFGGNGTSGQFLTRGMGQNSVDIKVDNAYSQILYXQ 112

QY 112 GETDMADFPDHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPKMP-- 169
Db 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDQAQ 153

QY 170 ---NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVLHTEGLY---RKSGDYAVPR- 219
Db 154 DLLKGLDKNWGRVRLNSFGAGNNGASYGASVFGKECP-----DGLFSYVRNDEKDYAGKG 209

QY 220 YRN---LKRLPDSP-----RRFANGQRAVLGWRKRFVR--RTYSRRDQVGLP 263
Db 210 FRNDNGGKTVFPYSALDKRSYLAKITGTFDGDHRIVLSHMKDQHRGIRTV---REBFAS 266

QY 264 AHSHEYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHN 323
Db 267 EKN-----SRITTKRQAPSYRE--TTQSNINLAYTKDGLFVEKLDANAV-- 310

QY 324 GKPMIDLRNKRYELAEWKQPPFPFEALRVHLNRNDYHDEKAGDAVENFFNQNTQARI 383
Db 311 -----LEKKYS--ADKQD-----NGVAGNVKGNHTRIATRSNMFNDFS 348

QY 384 ELRHQPIORLKGWQVYLGQ--KSSAL-----SATSEAVKQPMLLDNKVQHS 430
Db 349 RLAEQTL--LK--YGINYRHOEIKPQAFINSEFEIKKEKATNEKKKRENEKIAKAYR 404

QY 431 PFGVEQANW-----DNFTLEGGVRVEKQKASIRYDKALIDRENYKQPLDILGAH 480
Db 405 LTNPTKTDGAVIEAIEIDGFTLTGGLRYDRFKVKTGDK----- 445

QY 481 QOTARSPALSGNW---YFTPOHKLSTASHQ--ERLPSTQEP--LYAHGKH-----VATNTFEV 531
Db 446 --TVSSSLNFSFGVWQPREHWSFSASHNYAGRSPLRYDALQTHGKRGIIISADGT--- 500

QY 532 GNKLNKRSNNIELALGVEGDRWQYNLALYRNRFNRYIYAQTLNDG--RGPKSTEDDSEM 590
Db 501 -----KAERARNTGIFVNDGTFPAAN-----GSY--FRQTIKDALANPQNRHDSVAV 546

QY 591 K-----LVRYNQSGADFYGAEIYFKPTPRYRIGVSGD 624
Db 547 REAVNAGYIKNHGYELGASVYRTGGLTAKVGVSHSKSPREYDTHOKLLSANPEFGAQVGT 606

QY 625 YVRG---RLKNLPSLP---GR--EDAYGNRPFIACDDONAPRVPAPARLPHLKASLTDR 675
Db 607 WTASLAYRFXK--PNLEIGHRGYVQKAVGSILAAQKQDQKLENVVRQGFVN----- 659

QY 676 IDANLDYVRVFAQNKLAREYETTPGHMLNLANRYRTRYGEMWYVKAONLNLNQSYA 735
Db 660 -----DVFAWKPLGKDTLNL-----VNLNVN-----NVFDKFYYP 689

QY 736 HSFSLSDT--PQMRSGFTGVGVNKF 758
Db 690 HSQRWNTNLLPGVGRDVRIGVNYKF 713

```

RESULT 5

US-10-181-319-2

; Sequence 2, Application US/10181319

; Publication No. US20030135032A1

; GENERAL INFORMATION:

```

; APPLICANT: Lewis, Thomas A.
; APPLICANT: Paszczynski, Andrzej
; APPLICANT: Crawford, Ronald L.
; APPLICANT: Cortese, Marc S.
; APPLICANT: Sebat, Jonathan L.
; TITLE OF INVENTION: Compositions and Methods for Bioremediation
; FILE REFERENCE: IDAHL19440
; CURRENT APPLICATION NUMBER: US/10/181,319
; CURRENT FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: PCT/US01/02386
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/177,251
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 687
; TYPE: PRT
; ORGANISM: Pseudomonas stutzeri
; US-10-181-319-2

Query Match          4.9%; Score 198.5; DB 12; Length 687;
Best Local Similarity 21.4%; Pred. No. 1.2e-09;
Matches 170; Conservative 111; Mismatches 286; Indels 229; Gaps 43;

QY 7 KPVLISILLINT-----PLIAQA-----HETEQSVGLTETVVVVGKSRPRATSGLLTSTASD 58
Db 21 QPMWMTALICAFVPGPQLAPAAPGSAASPDSTTLPEITVTAEKIERP---LERVPASV 76

QY 59 KIISGDTLRKAVNLGDALDG--VPGIHASQVQ--GASAPVIRGQTGRIRKVLNHHGETGD 116
Db 77 AVIDGMDAESQSGITSLKQLEGRIPLGSFPQFQAQNSGPMVRGL-----TAN 123

QY 117 MADFSPDHAIMVD---TALSQ-----QVEILRGVPTLLYSSGNVAGLVADVADGKI 163
Db 124 FNSFSSSTLLLDVGVPTTLTAQFBSGMLDLDRIVIRGPQSTLYGNAEAGV--IAHSL 181

QY 164 P-EKMPENGVSSELGLR-----LSSGNLEKLTSGGI--NIGLKNFVLHTEGLYRKS 212
Db 182 PMDATPRASVSAAEAGSRNKRVMRFALSQLVLEERYLYGVSQGNWSSQDGDINTH--TGHKA 240

QY 213 GDVAVPRYRNALKRLPDSPPRFANGQHRVGLGWRKFRYRTYSDRRDOVGLPAHSHEYDDC 272
Db 241 DD-----REQNML--NLGLRWAPGATDVV---MRYAHOEYDDGASLWSPGAPRK--RVA 289

QY 273 HADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKPMIDURN 332
Db 290 SGTPSNRS-----EGQTLSENVQHEFASFASGLRLHLSVTANNEFKD 328

QY 333 KYRELAEWKQPPFGEALRVHLNRNDYHDEKAGDAVENFFNQNTONARIELRHQPIGR 392
Db 329 -----RIQDQTDMPADVL--HVGRD--HH-----LRTUSQEPFVE-----Q 362

QY 393 L-KGSN--GVQYLQKSSALSATSSEAVKQPMLLDNKVQHSYFFGVEQANWD-----NFTL 444
Db 363 LGEASMLAGV-YADRSNDLHSTKTMGM--LSDIRADQSDTAALFTHWNVPLSADWSI 419

QY 445 EGGVRYVEKQASIRYDKALIDRENYKQPLDILGAHROTARSFALSGNWFYTPQHKLSLT 504
Db 420 DAGARVERNEVQLR--PQGATSEKGTHTVSPRLALQHQ-----IT 458

QY 505 ASHQEELPSTOEIYAHGKHVATNTFEVGNKLNKRSNNIELALGYEG-----DRWOYNLA 560
Db 459 ANHQWYVSASRGVRYTGGFNVLPATL-----GYLPYDTEKNWSYETGLKGWLLDKRIRYSLA 514

QY 561 LYRNRFNGYIYAQTLNDGRGPKSIDDSMKL-----VRYNQSG--ADFYGAEIYFKPT 614
Db 515 AY-----LMD-----IDMQVMQMPVTGVMTYITSAATATSKGLELDV----- 551

QY 615 PRYRIGVSGDYVYG---RLKNLPS-----LPGREDAYGNRPFIADQ-----DQN 655
Db 552 -----DYLLGGWQKLGGLAWNHRFPDHRDGEADYDGNQNPFPAPDLTGHGLIRYD 602

```

100

Db 326 SHTLIPGHAQISYGAIVYRQOQKPEGSAT-----LYPEGNI DFTSLYFODEMTKSY 379

Qy 444 LBGQVVRVEXKASIRYDKALIDRENYIKOPLP---DLGHRQTARSFALSGNWTPOHK 500

Db 380 VNIIV-----GSRYDR-----YKGFNPRAGELKAERLSPRA-AIS-----VSPTDW 419

Qy 501 LSL--TASHQERLPSQELVYAHGKHVATWTFEYGN-----KHLNERSNNIELALG--- 549

Db 420 LMVYSISSAFRAPTWAEYRDDVHF-----YKGFKNYVWPNLNKPNNTREIGAGIQ 475

Qy 550 -----YEGDRWQYNLALYRNFRNGYIYAQTILNDGRGPKSIEDDSEMKLVRYNOSGADFYG 604

Db 476 LDGLLTNDRLQLKGVFGTDARNYIATRV-----DMKMRSYSYNVSRRATWG 524

RESULT 8

US-10-043-344-11

; Sequence 11, Application US/10043344

; Publication No. US2003008086A1

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Harkness, Robin E. B.

; APPLICANT: Schryvers, Anthony B.

; APPLICANT: Chong, Pele

; APPLICANT: Gray-Owen, Scott

; APPLICANT: Mordin, Andrew D.

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES

; FILE REFERENCE: 1038-1221 MIS

; CURRENT APPLICATION NUMBER: US/10/043,344

; PRIOR FILING DATE: 2002-07-01

; PRIOR FILING DATE: 1996-05-17

; NUMBER OF SEQ ID NOS: 160

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 11

; LENGTH: 914

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-10-043-344-11

Query Match 4.5%; Score 180.5; DB 15; Length 914;

Best Local Similarity 19.7%; Pred. No. 9.8e-08;

Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPLLAQAH-----ETEQSVG-----LETVTVVG-KSRP 44

Db 2 TKKPYPRLSIISCLLSYVKAETQSIKDTKEAISSEVDQSTEDSELETISVTAEKIRD 61

Qy 45 RATSGLLHTSTADKII--SGDITLROKAVNLGDALDGVPGIHASOYGGGASAPV-IRGQT 101

Db 62 KXNEV-----TGLKIIKTSISISREQVILNIRDLTRYDPGLSVVQGRGASSYIRGMD 117

Qy 102 GRRIKVLNHHGTGDMADFSPDHAIMVDYDTSALSOQVEILRGPVTL--YSGSNVAGLVDVA 159

Db 118 RNRY-----ALLVDGLPQTSYVQSVPLVARSYSGTGAINIEYIE 158

Qy 160 DGIKPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLG-KNFVHTEGLYR----- 210

Db 159 NVKAVE--ISKGSSEYNGNLAGSVTFQSKSAADILEGDKSWIGIQTKNAYSNNKGFTH 217

Qy 211 -----KSGDY---AVPRYNLKLPSDPRFANGQHR-AVLGWRKFRYRTYSDRRDQY 260

Db 218 SLAVAGKGGFDGVAITYRN-----SIEQVHKDALKGVS--YHRLIAKPEDQ- 265

Qy 261 GLPASHEDYDCHADITWQSL-INRYLQIYPHLLTEEDVDY-----DNP----- 305

Db 266 --SAYFMQDECPKDDYNSCLPFAKRPAILSSQRETVSVSDYTGANRIKPNMKEYESQS 323

Qy 306 -GLSCGFHDDDAHAHANGKFWDLNKRVELRAENKQPPGPFALRVHLNRNDYHDE 364

Db 324 WFLRGYHFSQ---HYIGG--IFEFTQKFDIR---DMTFPAY-----LRSTE 364

Qy 365 KAGDAVENFENNOTQA--RIELRHOPIGR-----LKGSMGVOVLGOKSSA 408

Db 365 KRDDSSGTFPKQDYQAYQRIE-----DGRGVNAYASGLYDEHHRKQRGVIEIYENKNK 419

Qy 409 LSATSEAV-----KQPMILLDNKVQH-----YSPFGVE----- 435

Db 420 AGIIDKAVLSANQQNIILDSYMQHTHCSLYPNFSPKNCRPTDRDKPYSYHSDRNVYKEKHN 479

Qy 436 -----QANW-----DNFT-----LEGGVRVEKOKASIRYDKALIDR 466

Db 480 MLQINLEKKIQONWLTHQIVFNLFDDFTSALQKHQVLTTRVTATAKIS--EKANETR 536

Qy 467 ENYKQ-----PLPDLG-----AHRQTARSPALSGNWTFTPOHKL 503

Db 537 RNYKKQPYLPKPTVGVVQDCHDYKGNSSNYRDCVKRLIKGNVYFAARNNNALCKYV 596

Qy 504 -----TASHQE-----RIPSTQEL 517

Db 597 DLGLGIRYDVSRTKANESTISVGKFKNFSWNTGIVIKPTBWLDSLVSRLSTGFRNPSFAEM 656

Qy 518 YAHGKHVATNTFEVGNKHLNKNERSNNIELALGYEGDRWQYNLALYRNFRNGYI-YAOTLN 576

Db 657 Y--GWRYGNNSEVYVGVKFKPETSBNQEPGLALKGDFGNIESHFSNAYRNLIATAFEBLN 714

Qy 577 ---DGRGPKSIEDDSEMKLVRYN--OSGADFYGABGEIYF----- 611

Db 715 KNGTGKANYGHNAQNAKLVGNITAQDLFNLGWLKRIPIGYWYATFAYNRVVKVDQKINAG 774

Qy 612 -----KPTPRYRIGVSGD-----YVRGLKNLPSLPQREDAYGNRPF 648

Db 775 LASVSVLFDALQFS--RYIIGLYDHSNTWGTINTMETQSKAKSONELLGKR-ALGNN-- 830

Qy 649 IAQDDONAPRVPAARLGFHLKASLTDRIDANLDYRVFAQNKARYETRETEGHMLNIGA 708

Db 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYVMAK-----NIMRLGI 865

Qy 709 NYRNTRYGENWYVYK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGNNVVF 758

Db 866 YNLFNRYRYTWEAVRQTAQGAVNQHONVGSYTRYAAS-----GRNYITLTLEMKF 914

RESULT 9

US-10-238-075-263

; Sequence 263, Application US/10238075

; Publication No. US20030148324A1

; GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238,075

; PRIOR FILING DATE: 2002-09-10

; PRIOR FILING DATE: 2000-03-10

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 263

; LENGTH: 713

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-238-075-263

Query Match 4.4%; Score 177.5; DB 12; Length 713;

Best Local Similarity 19.5%; Pred. No. 1.3e-07;

Matches 167; Conservative 103; Mismatches 283; Indels 305; Gaps 41;

Qy 6 LKPIVLSILLINTPLLAQAHETEQSVGLE-----TTVVVGKSRPRATGSL--LHTSTA--- 56

Db 11 LKGRALFSLFAAPMI---HATDSVTTKDGETITVTADANTATEATDGYQLSTATLT 67

Qy 57 -----SDKIISGDTLRQKAVNLCDALDGVPG-IHASOYCGGASAPVIRGQT 101

Db 68 DMPMLDIPQVNVNVSQVLE-----NQNATTLDEALYNSVNVQNTLGGTQDADFVRRGFG 123

```
102 GRIKVLNHHGCTGDMADSPDPAHVD-----TALSQOQVEILLRGVPTLLYSNGN 151
124 ANR-----DGSINTWGLRTVLPSPNAATERVEVLKGPASTLYGILD 165
152 VAGLVADVADGKPIKMPENGVSGLGLSSGNLEKLTSGGINIGLKG----NFVLHTEG 207
166 PGLINVT-KRPEKTFHGSVA-----TSSFGGTGQLDITGPIEGTQL 210
208 LYRKSGDVAVPYRNKRLKRLPDSPPRANGQRAVLGWRKFRFYRYSDRDOYGLPAHSH 267
211 AYRLTCEV-----QDEDYWRNFGKERSTFIAPS----- 238
268 EYDDCHADIIWQSLINKRYLQLYPH-----LITBEDVD-----YDNFGL 307
239 -----LTFVGD-NATVTLMSHRDYKTFPDRGTIFDLTKQPVNVDRKIRDEF-- 286
308 SCGFHDDDAHAHANGKPIDLRNRYELRAEWKPPGPFALRVHLNRNDYHDEKAG 367
287 ---FNITD-----GQSDLAQLNAEYHLNSQWTARF-----DYSYSQ--- 319
368 DAVENFNQ-----TONARIEL-----RHQPIGRKLGSGVQVYLGOKSSAL 409
320 ---DKYSDQARVAYDATGTLTRVDATQGTQRMHATRADLOGN--VDIAGFYNEIL 374
410 SATS-----EAVKQPMILLNKNVQHYSPFG-----VEQANW----- 439
375 GGVSYEYDILRTDMTRCKKAKDFNIYNPVYNTSKCTVTSASDSQTIKQENYSAQAQD 434
440 -----DNFTLEGVRVEKOKASIRYDKAL-IDRENYKQPLDPLGAHRTQARSFALSGNW 493
435 ALVLTNDWIAVAGRIQYTYTAGKGRPFNVWVDSRDEQWTFGLGVLYKLTFSVSLFANY 494
494 --YFTPOHKLSTASHQERLPSTQELYAHGKHVATNTFEVGNKHLNKRNNIELALGYE 551
495 SQTFFMQ---SSIASYIGDLPE-----SSNAYEVGAK-----FELFDGIT 532
552 GDRWQYNLALYRNRFNGNYLAOTLND-----GR-GPKSIEDDSEMKLVRYNQSGADPY 603
533 AD-----IALFDIHKRNVLVTESIGDETAKTAGRVRSRGVEVDLAGALTEVINIIASVG 587
604 GAEGEYFFKPTPRYRIGVSGDYVRGRLKNLPSLPGRDAYSGRPFIAQDDQNAAPRPAAR 663
588 YTDKAVLEDP-----DYAGKPLNVPRHGT-----SLFLTYDIHMPGNNTIT 630
664 L--GFH--LKASLTDRIDANLDYRYVFAQNKLYARTPTGHHMLNLGANYRNRTRYGBW 719
631 FGGGHHGVSSRSATN-----GADYY-----LPGYFVADAFAPAAAYKMKLOQY-PV 671
720 NWYVKADNLNQSIVYAH 737
672 TLQNLVKNLFDKTYTYS 689

RESULT 10
US-10-043-344-5
; Sequence 5, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Mordin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 913
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
; US-10-043-344-5

Query Match
Best Local Similarity 4.0%; Score 163; DB 15; Length 913;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

Qy 5 TLKPI-----VLSILLINTPLLAQH---ETEQSVG-----LETVTVG-KSRP 44
Db 2 TKPYFRLSIISCLLSICYKAEITQSIKDTKEAISSEVDTSQSTEDSELETISVTAIKVRD 61
Cy 45 RAYSGLLHTSTASDKII--SGDTRLOKAVNLGALDQVPGIHASQVGGGASAPV-IRGQT 101
Db 62 RKNEV---TGJGLKIITKTSISRRQVNLIRDLTRYDFGISVVEQGRGASSYSIRGMD 117
Qy 102 GRIKVL-----NHHGETGDMADSPDPAHVDLTAQVMDTALSQOQVEILLRG 142
Db 118 RNRVALLVDGLPQTQSYVQSPVARSYSGTGAINIEIYENV-----KAVEISKG 169
Qy 143 VTLLYSSENAGLV-----DVADG-----KIPEKMPENGVSGLGLRLSSCNLE 186
Db 170 SSSYTGALAGSVTTFQSKAADILEDGKSWGITQKNAYSSKNKGFTHSLAVAGKQGGFE 229
Qy 187 KLTSGGINIGLKN-----FVLHTE--GLYRKSGDYAVPYRNKLR 225
Db 230 -----GVAIYTHRNSIETQVHKDALXGVSQSYDRFIATTEDQSAFYVMQDECLDGYDKCT 284
Qy 226 LPDSPPRANGQRAVL-----GWRKFRFYRYSDRDOY--GLPA 264
Db 285 SPKRPATLSTQRETQSVSDYTGANRIKPNMPKYESOSW---FLRGGYHFSEQHYIGI-- 339
Qy 265 HSHEYDDCHADIIWQSLINKRYLQLYPHLLTBEEDVDYDNPGLSCGFHDDDAHAHANG 324
Db 340 --FEFTQCKFDI-----RQWTFPAYLRPTEDKDLQ---SRPPYKQDYGAOYHIG 384
Qy 325 KPWIDLNKRYELRAEWKQPPGPFALRVHLNRNDYHDEKAGDAVENFNNQTONARIE 384
Db 385 ---DGRGVKY-----ASGLYFDEHHRKQVRG--IEYIYENKKNAGIIL 422
Qy 385 LRHQPIGRKLGSGVQVYLGOKSSALSATSBAVKPMLLDNKNVH----- 428
Db 423 -----KAVLSANQONILDSYMRHETHCSLYPNFSKNCRPT 457
Qy 429 ---YSPFGVE-----QANW-----DNFT----- 443
Db 458 LDKPYSYHSDRNVYKKEKHNMLQLNLEKIQONMLTHQIAFNGLGDDFTSALQHKDYLTR 517
Qy 444 --LEGVYRVEKOKASIRYDKALIDRENYKQPLP-----DLGAHQRTARSFA----- 488
Db 518 RVIATASSISEKRGAEARRNGL---OSSPYLYPTPKAELVGGDLNCGYQKGSNSYDCKVRL 574
Qy 489 LSG-NWYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTFEVGNKH-LNKERSNNIEL 546
Db 575 IKGNKYTFAARNMAL-----GKYV---DLGLGMYDYVSRVTKANESTI 614
Qy 547 ALG-YEGDRWQYNLALYRNRFNGNYIYAQTLNDG-RGPKSIEDDSEMKLVRYNQSGADPY 604
Db 615 SVGRKQKPSMNTGIVIKPTWELDSL--RLSTGPRNFPSP---AEMYGMWYGGKDTDVI 668
Qy 605 AEGEYIFKP-----TPRVIRIGVSGDYVRGRLKNLPSLPGRDAYSGRPFIAQDDQNAAPRV 660
Db 669 CK----FKPSTRNQEPGLAKGDFGNIESHF-----SNAYNLIAFAEELSKNCTTG 718
Qy 661 AARLGTFH-----LKASLTDRIDAN-----LDYRYVFAQNKLA----- 692
Db 719 KGNVGYHNAQNAKLVGNVITQAQLDFNLGWLKRIPIGYWATPAYNRVVKVDQKINAGLASVS 778
Qy 693 --RYETRTPGHHMLNLGANYRNRTRYGEWNYVKADNLNQSIVYAHSSFSLSDFQWGRSF 750
```

```

Db 779 SYLFDALQPSRYIIGLYDHPSTNT-----WGIK-----TMFTOSKAKSONELLGKRA 825
Qy 751 TG--GVNVK 757
Db 826 LGNNSRNVK 834

RESULT 11
US-10-043-344-7
; Sequence 7, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murdin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 912
; TYPE: PR1
; ORGANISM: Haemophilus influenzae
US-10-043-344-7

Query Match 4.0%; Score 161.5; DB 15; Length 912;
Best Local Similarity 19.6%; Pred. No. 6.5e-06;
Matches 200; Conservative 129; Mismatches 315; Indels 377; Gaps 54;

Qy 5 TLKPI-----VLSILLINTPLAQAH---ETEQSVG-----LETVTVWG-KSRP 44
Db 2 TKPYFRLSIISCLLSICVYKAEQTSIKDTKEAISSEVDQTSTEDSELETISVTAEKIRD 61
Qy 45 RATSGLLHTSTADSKII--SGDTLRQKAVNLGDALDVGPGIHAQYGGGASAPV--IRGQT 101
Db 62 RKDNV-----TGLGKIITSEISREQVNLRLTRYDPLGSLVVEQGRGASSGVSIRGMD 117
Qy 102 GRRIKVL-----NHHGTGDMADSPDHAIMVDLTALSQOQVEILRGP 142
Db 118 RNRVALLVDGLPQTQSYVVSPLVARSGYSGTGAINETIENV-----KAVEISKGG 169
Qy 143 VTLLYSSGNVAGLV-----DVADG-----KIPKMPENGVSCELGRLSSGNLE 186
Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSNNKNGFTHSLAVAGKQGGFE 229
Qy 187 KLTSQGINIGLKNFV---LHTEGL-----YRKSGDY----- 215
Db 230 -----GLAIYQTSNIETQVHKDALKGVQSDRLIATTDKSSGYFVIQGECPNGDDKCA 284
Qy 216 -----AVERYNLKRLPDSRRFAN-----GQH---RAVLGWRKRFYRRT 252
Db 285 KPPATLSTQSTVSDVTGANIKPNPKYESQWFLRGYHFSHQHYIGIFETQOQ 344
Qy 253 Y-----SDRRD-----QYGLPAHSHE-----YDQCH----- 273
Db 345 FDIRDMTFPAPLSPATERDDSRFPYMQDHGAYQHIEDGRGVKYASGLYFDEHHRQVR 404
Qy 274 -----ADII-----WOKSLINKVLO-----LYPHLLTEEDVDYDNPGLSCG 310
Db 405 GIEYIYENKAKGIDKAVLSANQQNIILDSVTRHTHCSLYP-----NPSKNCR 453
Qy 311 -----FHDDDDAHAAHNGKFWIDLRNKKRYELRAEW---KQPPP-GFEALRVHLNR 357
Db 454 PTLDPKPSYYSRDRNVYKEKN---MLQL-NLEKKIQQWLTHQIVNFGDFFDFTSALQH 509

```

```

Qy 358 NDYHHEKAGDAVENFFNNQTONARIELRHQPIRLKSGWGVQVYLGOKSALSATSBAVK 417
Db 510 KDY-----LTFRVATATADSIPRKPGETGKPRNGLOS----- 540
Qy 418 QPMLLDNKKVQHYSPFFGVEQANWDFNFTLEGGRVVEKQKASIRYDKA---LIDRENY--- 470
Db 541 QPYLYP-KPEPY-FAGQDHCNY-----QSSSNRYRDKVRLIKGNKYFAAR 585
Qy 471 ----KQPLPDLG-----AHROTA-----RSFALSGNMYFTPOHKLSLT--ASHQ 508
Db 586 NMALGKYVDLGLGIRVDVSRTKANESTISVGKFNFSWNTGIVIKTEWLDLSYRLSTG 645
Qy 509 ERLFSTQELYAHGXHVATNTFEVGNKHLNKNERSNNIELALGYBODRQOYNALYRNFEN 568
Db 646 FRNPSFSEMY--GWRYGGKNDVVYGVFKPETSQRNQEFGLAKGDFGNIEISHFSNAYRN 703
Qy 569 YI-YAQTIN---DGRGPKSIEDDSEMKLVYN--OSGADFYG-----AEGEYI 610
Db 704 LIAFABELSKNGTGKGNVGYHNAQNAKLVGNVITIAQLDFNGLWKRIPIYGVWATPAYNOVK 763
Qy 611 PK-----PTPRYRIGVSGD-----YVRGRKLNPLSPFGR 640
Db 764 VKDQKINAGLASVSYLFDALQPSRYIIGLYDHPSTNTWGTMTQSKAKSONELLGKR 823
Qy 641 DAYGNRPFIADDDQNAVRPAARLGFHLKASLTDRIDANLDYRVFAQNKLYARTETPG 700
Db 824 -ALGNN---SRDVKSTRKLTRA---WHI-----LDVSGYVANK----- 855
Qy 701 HHMLNLGANYRNTRYGEWNVVK-ADNLLN--OSVVAHSSFLSDTPQMGSRFTGGVNVK 757
Db 856 NIMRLGIYNLFNYEYVTVWEAVRQTAQAVNQHQNVGSYTRYAAS---GRNYTLTLEMK 911
Qy 758 F 758
Db 912 F 912

RESULT 12
US-10-043-344-9
; Sequence 9, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murdin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 912
; TYPE: PR1
; ORGANISM: Haemophilus influenzae
US-10-043-344-9

Query Match 4.0%; Score 161.5; DB 15; Length 912;
Best Local Similarity 19.6%; Pred. No. 6.5e-06;
Matches 200; Conservative 129; Mismatches 315; Indels 377; Gaps 54;

Qy 5 TLKPI-----VLSILLINTPLAQAH---ETEQSVG-----LETVTVWG-KSRP 44
Db 2 TKPYFRLSIISCLLSICVYKAEQTSIKDTKEAISSEVDQTSTEDSELETISVTAEKIRD 61
Qy 45 RATSGLLHTSTADSKII--SGDTLRQKAVNLGDALDVGPGIHAQYGGGASAPV--IRGQT 101
Db 62 RKDNV-----TGLGKIITSEISREQVNLRLTRYDPLGSLVVEQGRGASSGVSIRGMD 117
Qy 102 GRRIKVL-----NHHGTGDMADSPDHAIMVDLTALSQOQVEILRGP 142
Db 118 RNRVALLVDGLPQTQSYVVSPLVARSGYSGTGAINETIENV-----KAVEISKGG 169
Qy 143 VTLLYSSGNVAGLV-----DVADG-----KIPKMPENGVSCELGRLSSGNLE 186
Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSNNKNGFTHSLAVAGKQGGFE 229
Qy 187 KLTSQGINIGLKNFV---LHTEGL-----YRKSGDY----- 215
Db 230 -----GLAIYQTSNIETQVHKDALKGVQSDRLIATTDKSSGYFVIQGECPNGDDKCA 284
Qy 216 -----AVERYNLKRLPDSRRFAN-----GQH---RAVLGWRKRFYRRT 252
Db 285 KPPATLSTQSTVSDVTGANIKPNPKYESQWFLRGYHFSHQHYIGIFETQOQ 344
Qy 253 Y-----SDRRD-----QYGLPAHSHE-----YDQCH----- 273
Db 345 FDIRDMTFPAPLSPATERDDSRFPYMQDHGAYQHIEDGRGVKYASGLYFDEHHRQVR 404
Qy 274 -----ADII-----WOKSLINKVLO-----LYPHLLTEEDVDYDNPGLSCG 310
Db 405 GIEYIYENKAKGIDKAVLSANQQNIILDSVTRHTHCSLYP-----NPSKNCR 453
Qy 311 -----FHDDDDAHAAHNGKFWIDLRNKKRYELRAEW---KQPPP-GFEALRVHLNR 357
Db 454 PTLDPKPSYYSRDRNVYKEKN---MLQL-NLEKKIQQWLTHQIVNFGDFFDFTSALQH 509

```

```

Db 62 RKDNEV---TGLGKIILKTSIESREQLNIRDLTRYDPGISVWEQGRGASSGYSIRGMD 117
QY 102 GRIRKVL-----NHGETGDMADFPDHAJWDTALSOQVILRGP 142
Db 118 RNRVALLVDGLPQTQSYVVSQFLVARSQSGTGAINIEYENV-----KAVISKGG 169
QY 143 VTLLYSSGNVAGLV-----DVAAG-----KIPEKMPENGVSGLGLRLSSGNLE 186
Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSSKNKGFTTHSLAVAGQGGFE 229
QY 187 KUTSGGINLGNFV---LHTEGL-----YRSGDY----- 215
Db 230 -----GLAIYTORNSIETQVHKDALGVQSYDLRIATTDKSSGYFVIOGCECPNGDKCAA 284
QY 216 -----AVPRYRNLKRLPDSRRFAN-----GOH-----RAVLGWRKRYFRRT 252
Db 285 KPATILSTOSETVSVDYTGANKIKENPMKYESQSWFLGGYHFSQHVIGIGIFETQOK 344
QY 253 Y-----SDRRD-----OYGLPAHSHE-----YDCH----- 273
Db 345 FDIRDMTFPAYSPTERRDSSRSFYPMQDHGAYQHIEDGRGVKYASGLYFDEHHRKQV 404
QY 274 -----ADII-----WOKSLINKRYLO-----LYPHLLTEEDVDYDNLGLSCG 310
Db 405 GIEYIYENKAGIIDKAVLSANQNIILDSYWRHTHCSLYP-----NFSKNCR 453
QY 311 -----PHDDDDAHAAHNGKPIDLNRKRYELRAEW---KQPPP-GFEALRVHLNR 357
Db 454 PTLDPKYSYVRSRNVYKEKHN---MLQL-NLEKKIQNNLTHQIVFNLFDFDFTSALQH 509
QY 358 NYVHDEKAGDAVENFNQOTQNAIELRHQPIGRLLKSGWQVYLGQKSSALSATSEAVK 417
Db 510 KDY-----LTRRVIAATADSIPRKEGETKPRNGLOS----- 540
QY 418 QPMLLDNKVQHSYFFGVEQANWMDNFTLEGVRVVEKQKASIRYDKA---LIDRENY--- 470
Db 541 QPVLVP-KPEPY-FAGODHCNY-----QSSSNYRDCKVRLIKKNYFYFAAR 585
QY 471 -----KOPLPDLG-----AHRQTA-----RSFALSGNNYFPPQHKLSLT--ASHQ 508
Db 586 NNMALGKYVDLGLIGIRYDSRTANESTISVGKFKNFSWNTGIVIKPTWLDLSYRLSTG 645
QY 509 ERLPSTOELYAHKHKVATNTFEYGNKHLNKNERSNNIELALGVEGDRWQVNLALYRNFEN 568
Db 646 FRNPSSEMY---GWRVGGKNDVYVVGKFKPETSROQFGLALKGDRGNIEIHSFNAYEN 703
QY 569 YI-YAOTLN---DGRGPKSIEDDSEMKLVRYN-QSGADFYG-----AEGEY 610
Db 704 LIAFABELSKNGTKGNYGYHNAQNAKLVGNITAOIDFNLGWLKRIPIYGYWATFAYNQVK 763
QY 611 FK-----PTPRYRIGVSGD-----YVRGRKLNLPSPGFE 640
Db 764 VKDQKINAGLASVSVLFDIAQPSRYIIGLYDHPNTWGINMTFTQSKAKSQNELLGR 823
QY 641 DAYGNRPFIADQDQNAFRVPAARLGFHLKASLTDRIDANLDYRVPFAQNKARYETRTPG 700
Db 824 -ALGNN---SRDVKSTKLTRA---WHI-----LDVSGYTWANK----- 855
QY 701 HMLNLGANYRNRTRGEEWNYVK-ADNLLN---QSVYAHSSFLSDTPQMGRSFTGGVNVK 757
Db 856 NIMRLGILNLFNRYRYTWEAVRQTAQGVNQHNVGSYTRYAAS-----GRNYTLLEMK 911
QY 758 F 758
Db 912 F 912

```

RESULT 13

```

US-10-238-075-721
; Sequence 721, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.

```

RESULT 14

```

US-10-098-808-2
; Sequence 2, Application US/10098808
; Publication No. US20030007981A1
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Rioux, Clement
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
; TITLE OF INVENTION: TRANSFERRIN-BINDING PROTEINS
; FILE REFERENCE: 9000-0049.20
; CURRENT APPLICATION NUMBER: US/10/098.808
; CURRENT FILING DATE: 2002-03-13

```

```

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol
; TITLE OF INVENTION: B.coli, and biological uses of these polynucleotides and of th
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238.075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 721
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-721

```

Query Match 4.0%; Score 160; DB 12; Length 540;

Best Local Similarity 22.1%; Pred. No. 3.9e-06; Matches 122; Conservative 78; Mismatches 229; Indels 124; Gaps 31;

```

QY 209 YRKSQDYAVPRYRMLK--RLPDSRRFANGQRAVLGWRKFRYRTYSDDRDQYGLPAHS 266
Db 27 YRINGSYDQGNRDTPDGRLENT--NYRNNSQGVWLGYNCGNHR--FGLSLDRYRLATOT 82
QY 267 HEYDDCHADITWOKSLINKRYLQLYPHLLTEE-----DVDYDNFGLSCGFHDDDAHAHA 321
Db 83 Y-YEDPDGSY-----EAFSVKIPKLEREKVGVDYDQVGYLK-----KIHFDA 126
QY 322 HNGKPIWDLNRKRYELRAEWKQPPFGEALRVHLNRNDYHDEKAGDAV--ENFFNNQQTQ 379
Db 127 YEQTIQRFAN--EVKTPQVPSEMIQALTVH-NKTD--THDKQYQAVTLQSHFSLPAN 181
QY 380 NARIELRHQPIGRLLKSGWQVY---LGQKSSALSATSEAVKQPMLLDNKQVHYSFFGVE 435
Db 182 NELVT-----GAQYKQDRVSRQSGGTTSSKSLTG--PINKETRTSRSYESE 225
QY 436 QA-----NW---DNFTLEGVVRVEKQKASIRYDKALIDRENNYKQPLDLAGHROTA 484
Db 226 QSTVSLFAQNDWRTFADHWITWTVGR--QYWLSSKLTGR--DGVSTAGIISDTSLARESA 281
QY 485 --RSPALSGNNYFPPQHKLSLTASHOER--LPSTOELYAHGKHVATNTFEVGNKHLNKR 540
Db 282 SDHEMVTSTLSRYSGFDNLELRAAPQAGYVFTLSQLFMQTSAGGSVTY--GNPDLKAEH 339
QY 541 SNNIELALGYEGDRWQVNLALYRNFYNYIYAOTLNDGRGPKSIEDDSEMKLVRY--NQS 598
Db 340 SNNFELGARYNGNQWLIDSAVYSEAKDYI--ASLICDGSIVCGNGTNSRSRYSYYDND 398
QY 599 GADFYGARGEITYFKPTPRYRIGVSGDYVVRGRL-----KNLPSILPGR----- 639
Db 399 RAKTWGLE-----ISAENYNGWVSPYISGNLIRROYETSLTKTNTGEPAINRIGLKH 452
QY 640 -----BDAYGNRRPFIADQDQNA--PRVPA-ARLGFHLKASLTDRIDANLDYRVRV- 685
Db 453 TLVMQANIIISDVFTRAASSAKDDSGTGTNVPQWATLNFVAVNTEF-----GNEDQYRIN 507
QY 686 FAQNKLA--RYET 696
Db 508 LALNNLTDKRYRT 520

```



```

Db      795 ILGRIYWGEGISDRUPBEGLYSLAYNRINIKARKLHNDFTNVSEPTLEAVQPGRIIA----- 850
QY      647 PFIAODQONAPRVPAAIRLGPHL-----KASLTDRIDANLDYRVYPAQN---KLARYETRT 698
Db      851 -SIGYDD-----PEGRWGLNLSGTYSQAKORDEVGGE---KVFGGSGIKRTINSKRT 899
QY      699 PGHMLNLGANYR-----RNTRYGEW-NWYVKADLLNQSVVYAHSSFLSDT 743
Db      900 RAWIYDYLTYATYTWKEFTLRAGIYNLTNRKYSTWESVRQSAANAVQDLGTRSAFA-- 957
QY      744 PQMGRSFTGGVNVKF 758
Db      958 -ARGRNFVTSMBMKF 971

RESULT 15
US-10-238-075-1077
; Sequence 1077, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are iso
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1077
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-1077

Query Match          3.9%; Score 155.5; DB 12; Length 682;
Best Local Similarity 18.9%; Pred.No.3.5e-05;
Matches 143; Conservative 103; Mismatches 22; Indels 289; Gaps 40;

QY      51 LHTSTASDKIIISGDTLRQKAVNLGDALDGVPGIHAISOYGGGASAP--VIRGQTGRRIKVL 108
Db      44 LRNPASVSVITSEQLOKKPVS--DLVDAVKDVEGISITGNGEKRPDISIRGLSGDYTLIL 101
QY      109 -----NHHGETGDMADFPD-HAIMVDTALSOQVELLRGPVTLILYSSGNVAGLV 156
Db      102 VDGRRSGRESRPNGSGGPEAFIPVPEAI-----ERIEVIRGPMSELGYSDAIGGVI 154
QY      157 DVADGKIPEK-----MPENG-----VSGELGLRISGG-NLEKL 188
Db      155 NIITKPVNQTDGVLGLGGIIQEHGKFGNSTTDFYLSGPLIKDKLGLQLYGGMYRKE 214
QY      189 TSGGING-----IGKPFVLHT-----EGL--YRKSQDYA 216
Db      215 DS--ISQGTAPAKDNKNITATQFTTESQKFVFEYGNQNVHILTPGSLDAWTMRG--- 269
QY      217 VPRYRLKELPDSPRRFANGQRAVLGW-----RKRFYRRTYSRRDOYGL 262
Db      270 -----NLXQ-PNSKRETHNSRSHWAAWNAQGEILLPEIAYVQEKVIREVSKGKKYN- 322
QY      263 PAHSHBYDCHADITWQ-----KSLNKRYQLQYPHLLTEEDVDYDNPGLSGCFHDDD 315
Db      323 -----HWDLNYESRKPEITTIIDAKVTAFLP-----ENV-----LTIG----- 356
QY      316 DAHAHAHGKPMIDLNRKNRYELRASWKQFPFGFEALRVHLNRNDYVHDEKAGDAVENFFN 375
Db      357 -----GQFQHAELRDDSATGKKT 374

QY      376 NOTQNAIRIELRHQPIGRLKGSGVQYLGQKSALSATSSEAVKQPMLLDNKVQHSFFGVE 435
Db      375 TETQSVSIKQKAVFIEN-----EYAATDLSLATG-----GLRLDN-----HEIVG-- 414

```


QY 436 QANW-----DNFTLEGGVVEKOKASIRYDKALIDRENYKQPLDLAGHROTA 484
 Db 415 -SWNPLRYAVYMLTDLNLKGLGIAKAFRAPS-----EVSFGTILTOGG 460
 QY 485 RSPALSGNWYFTPO---HKLSTASHOERLPSTOBYLAHGKHVANTFEVGNK-----534
 Db 461 ASI-MYGNRDLKPETSVERIGIYNSDGSASATLFTNDFKNLTSYDICTKDPVTGL 519
 QY 535 -----HLNERSNNTIELALGYE-GDRQVYNLALYRNFGNYIYAOTLNDGRGPKSIED 587
 Db 520 NTFYDNGVANGINRGVLAQIPYDKHVS-----ANY-----TFDSR--RKSDD 565
 QY 588 SEMKLVRYNSGADFYGAEGEIVFKPTPRYRIG--VSGDYVRGLKNLPSLPGRDAYGN 645
 Db 566 S-----LNGSKLKEP-----LERTPRHAANAKLEWDYTD-ITFYSSL-----NYTK 608
 QY 646 RPFIAODDQAPRPAARLGFHLKASLTDRIDANLDY 682
 Db 609 QIWAQAQ--RNGAKVPRVRNGF-----TSMIDGLNY 636
 RESULT 16
 US-10-043-344-94
 ; Sequence 94, Application US/10043344
 ; Publication No. US2003008086A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Harkness, Robin E.
 ; APPLICANT: Schryvers, Anthony B.
 ; APPLICANT: Chong, Pelle
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Murdin, Andrew D.
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
 ; FILE REFERENCE: 1038-1221 MIS
 ; CURRENT APPLICATION NUMBER: US/10/043,344
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 08/649,518
 ; PRIOR FILING DATE: 1996-05-17
 ; NUMBER OF SEQ ID NOS: 160
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 94
 ; LENGTH: 908
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-043-344-94

Query Match 3.8%; Score 155; DB 15; Length 908;
 Best Local Similarity 19.3%; Pred. No. 2.7e-05;
 Matches 187; Conservative 122; Mismatches 360; Indels 300; Gaps 47;
 QY 14 LLINTPLLAQAHETEQSVGLTETVWVGKSRPRATSGLLHTSTASDKIISGDTL-ROKAVN 72
 Db 16 LMTALPYVAENVQAEQAKOLDTIOVKAKQKTRDNEVTGLKLVKSSDTLSKEQVLN 75
 QY 73 LGDALDGVPGIHASQYGGASAPV-IRGQTRRIKV-----LNHHG 112
 Db 76 IRDLTRVDPGIADVQEGSGASSYIRGMDKNRVSLTVGVQSQSYTAQAALGGTGTAG 135
 QY 113 ETGMDAFSPHAIMVDTALTSQQEILRGPTLYSSGNAGLV-----VADGK- 162
 Db 136 SSGHAINLEYENV-----KAVEISGNSSEYGNALAGSVAFQTKTAADIIGGKQ 187
 QY 163 --IPKMPENG-----VSGELGLRLSSGNLEKL-----TSGGINIGLGRNFV 202
 Db 188 WGIQSKTAYSGKDHALTQSLAGRSGGAELLYTKRGREIHAHKDAGKGVGSFNRV 247
 QY 203 LHTGLYKSG-----DYAVPR-----YENLRKP 227
 Db 248 LDED---KKEGGSQRYFIVEECHNGYAAACKNLKEDASVYDKRXTVSTQDYTGSRLL 304
 QY 228 DSPRRFANGQRAVLGWR--KRFFY-----RRTYS--DRRQYGLPAH--SHEY-----269

Db 305 ANPLEYQSQWLFPRGWHLDNRHYVGVAVLERTQOTFTDRM-TVPAYFTSEDDYVPGSLKG 363
 QY 270 -----DOCHAD-----ILWOKSLINKYLLQYLPHLLTEED-----299
 Db 364 LKYSQGNKABRLFVQGGSTLQIGYGTGVFYDERHTKNRYGYEYVHNADKOTWADYA 423
 QY 300 -VDYDNPGLS-----CGFHDDDDAHAAHANGKPMIDLNRK--YE-----LRAEWK 342
 Db 424 RLSYDRQGI DLNRLQOQTHCS-HDGSCKNCRPDGKNKPYSFYKSDRMIVYEEARNLPQAVFK 482
 QY 343 QPPP-----GFEALRVHLNRNDYH-HDEKAGDAV-----ENFENNQTON-----380
 Db 483 KAFDTAKIRHLSINLNGYDRFKSQLSHSDYTLQNAVQAYDLITPKKPPFPNGSKNDPVRV 542
 QY 381 -ARIELHQPTIGRLKSGWQVYLQKSSALSATSEAVKQPMLLDNKVQHYVFFGVEQAN 438
 Db 543 SIGKTTVNTSPICRGNN-----TYTDCPTRNIGGNGYAAV 579
 QY 439 WDNFTLEGGVREKOKASIRYDKALIDRENYKQPLDLAGHROTARSFALSGN-----492
 Db 580 QDNVELG---RWADVAGIRYDY---RSTHSEDKSVSTGTHRN-----LSWAGVVLK 626
 QY 493 ---WYFTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLNKRSSNNIELALG 549
 Db 627 PFTWM-----DLTYRSTGCFRLPSPAENYWRAGESLXTLD-----LXPEKSFREAGIV 676
 QY 550 YEGDRWQYNLALYRNFGNYI---YAQTLNDR---GPKSIEDSEMKLRYNQSQ-AD 601
 Db 677 PKGDFGNLEASVFNAYRDLIAFGYETRTQNGTSASGDPGYRNAQARIAGINILGKID 736
 QY 602 FYGAEGEIVFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFTAQ--DQONAPRV 659
 Db 737 WHGVWGGI---PDGLYS---TLAYNRIRVKD-----ADTRADRTFVTSYLFDAVQPSR 783
 QY 660 PAARLGF-----HLKASLTDR-----IDANLDYRVFAQNKLAFTYETRT 698
 Db 784 YVLGLGYDHPDGIWGINTMTFTYSKASVDELGSQLINGNANAKAASRRTPWYVTDV 843
 QY 699 PG-----HHMLNLCANYRNTRYGEW-NWYVKADNLNQ--SVYAHSSFLSDTPQMGSR 749
 Db 844 SGYVNIKKHLTLRAGVYNLLRYVYVWENVRQTAGGAYVQHNKGVNRYAAP-----GRN 899
 QY 750 FTGGVNVKF 758
 Db 900 YTFSLMKF 908
 RESULT 17
 US-09-815-242-5085
 ; Sequence 5085, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haseibeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Cart, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5085
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5085

Query Match          3.8%; Score 154; DB 9; Length 813;
Best Local Similarity 20.0%; Pred. No. 2.8e-05;
Matches 162; Conservative 101; Mismatches 265; Indels 284; Gaps 43;

QY 47 TSGLLHTSTASDKIISG---DTRQKAVNLGDAL-DGVPGIHAS-----QYGGGASAPVI 97
Db 131 TSSDLVSATRQSTVIEHAQLBELQSGDSLATVLAKAVPGMSDSSRTITEY-----181
QY 98 RGQT--GRIKVLNHGSETGMADFPDHALMVDLTALSOQVEILRGVPTLLYSSGNVAGL 155
Db 182 -GOTLRGKSMVMVDGVPNTNRSSRNLA-NIDPALIERIEVIRGS-SAIYSGATGGI 238
QY 156 VDV---ADGKIPKMPENGVSGBGL-RISS-----GNLEKLTSGGINIGLKNF---VL 203
Db 239 ISITRPAAG--ENRAETRLSATSPLTRGSDGLGGFQFYFAGSLG-ALDYSFDRGTR 294
QY 204 HTEGLYRKSGDYAVPRYNLKRLPDSPRRFANGQHR--AVLGWRKRYRRTYSDRRDQYG 261
Db 295 HVGASYDAHGRIAPE-----ESQGLDFSNVYNTGGKLGRL-----331
QY 262 LPAHSHEVDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDN-----PGLSCGFHD 313
Db 332 -----IDE-----NQEVOLALSHYDARQOTDYATDPRVARLPFG-----365
QY 314 DDDAHAAHNGKPIDLANKYEIRAEWKQFPFGFEALRVHLNRNDVHHD-----363
Db 366 --SVPAKAIKLE-IDEQNRINTIANLE--YENLDILGSRLSAQLYRDYFTRFTPFDA 420
QY 364 ---EKAGDAVENFNNTQNTARLHOPITRLKGS---WGVOLGOKS-----S 407
Db 421 RAVSTRGQNVQIQMSVFGSRLTLR--TFLGESNTLWGGDYNGERSDMPDLVDVDP 479
QY 408 ALSATSEAV-----KQPMLLDNKVQHYSPFQVEQANWD-NFTLEGGVVRVKKQA-----455
Db 480 AYDASGGLVFDKIGKLTYPPLRTSRAGAFQAQLQHRFDEHWSIDGGLRYEYSTAEFDFFI 539
QY 456 -----SIRYDKALID-----RENY--YKQ--PLPDLGAHROTAR 485
Db 540 PLSKSAASPVTVKGGDLDYDAVLSNLGIVYSPVAGQEIYASFQGFQDPDVGQLKRNAR 599
QY 486 SPALSGNNYFTPOHKLSTASHOERLPSTQBELYAHGKHVATNTFEVGNKHNKERSNIE 545
Db 600 -----RGFDIGSSNLEPVKNTNYYE 618
QY 546 LA-LGVEGRWOYNLALY--BNRFGNYIYAOTLNDGRGPKSIEDDSEMKLVRYNQSGADF 602
Db 619 LGWRGAIQNTVIGSLALFTYTSKLGD---VQSFNNG-----LILFTRKER---I 661
QY 603 YGAGEEI-YFKPTPRYRIGVDYVRGLKNLPSLPGREDAYGNRPFIADQDNAPRVA 661
Db 662 YGVEASADWLSDEVWAGGASATWNRGREK-----PDGKD-----WQDMTGYRVPP 707
QY 662 ARLGFH-----IKASLTDRIDANDLYRVFAQNKARLET-----RTGHH 702
Db 708 LKLTAYLQKPDADWNNLQNTFFDPSKDYRLDGVESFGRRQVSTYTTVDLVSQYRITPDD 767
QY 703 MLNLGANYRRNTRYGEWNWYKADNLLNQSVY 734
Db 768 QLSLG-----IQNLFNRDIY 782
```

RESULT 18

```
US-10-043-344-107
; Sequence 107, Application US/10043344
; Publication No. US20030088086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Hartness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Murgin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERIN RECEPTOR GENES
; FILE REFERENCE: 1039-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-043-344-107

Query Match          3.8%; Score 154; DB 15; Length 911;
Best Local Similarity 19.4%; Pred. No. 3.4e-05;
Matches 196; Conservative 125; Mismatches 331; Indels 360; Gaps 50;

QY 5 TLKPI-----VLSILLINTPELLAAH---ETEQSVG-----LETVVVG-KSRP 44
Db 2 TKPYFRLSIISCLLISCVKRAETQSIKDTKEAISSEVDTQSTEDSELETISVTAEKIRD 61
QY 45 RATSGLLHTSTASDKII--SGDTLQKAVNLGDALDGVPGIHASQYGGGASAPV-IRGQT 101
Db 62 RKDNEV---TGLGKIITKTSISREQVNLNRDLTRYDPGISVVEVQGGGASGYSIRMD 117
QY 102 GRIKVLNHGSETGMADFPDHALMVDLTALSOQVEILRGVPTLL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPTQSVYVQSPVARSYSGTGAINIEIYE 158
QY 160 DOKIPEKMPENGVSGBGLRLSSGNLEKLTSGGINIGL-G-KNFVLTETGLR-----210
Db 159 NYKAVE-ISKGSSSEYGNALAGSVTFQSKSAADILEGDKSWGIGTQKNAYSSKNKGPTH 217
QY 211 -----KSGDY--AVPRYNLKRLPDSPRRFANGQHRVAVLGWRKRYR--RTYSDRRDQ 259
Db 218 SLAVACKQGFSGVVAITYORNSEE-----TOVHKDALKGVSQYERFIATTDKSSGY 268
QY 260 YGLPAHSHEYDD-CHADIIMQKSLINKRYLQLYPHLLTEEDVDY-----DNP-----305
Db 269 FVIQCECPNGDDKCAA-----KPPAKLSQSETVSVSDYTGANRIKPNPMKYESQ 318
QY 306 --GLSCGFHDDDDAHAAHNGKPIDLANKRYVELSAEWKQPPGPFQPEALRVHLNRNDYHHD 363
Db 319 SWFLRGYHFSQ---HYGG--IFEFTQKFDIR---DMTFPAY-----LRST 359
QY 364 EKAGDAVENFNNTQNTONA--RIELRHQPIGR-----LKSGWGVQYVGGQSKS 407
Db 360 EKRRDRTGTFPKQDYQAYQRIE---DGRGVNVASGLYFDEHHRKQVRGIEYIYENKN 414
QY 408 ALSATSEAV---KQPMLLDNKVQH-----YSPFGVE-----435
Db 415 KAGIIDKAVLSANQONIIILDSYMRTHCSLYPNPSKNCRPTLDKPYSYRSDRNVYKKEH 474
QY 436 -----QANW-----DNF-----TLGGVVRV 450
Db 475 NMLQLNLEKKIQNWLTHQIVFNLGDDFTSALQHKDYLTRRVATANIISGTVAGKERN 534
QY 451 EKQKASIRYDK-----ALIDRENY-----KQPLPDL 477
```

Db 535 GYEQPILYKPKVDFVQDCHYKSSNSYSDCKVRILKGNYYFAARNMNLGKYIDL 594
Qy 478 G-----RSPALSGNMYFTPOHKLST--ASHQERLPSTQELYA 519
Db 595 GLGIRVDVSTKANESTISVGKFKNSWNTGIVIKPTWELDLXYLSTGFRNPSFAEM- 653
Qy 520 HGKRVATNTFVGNKHLNERSNNIELALGYGDRQWNLALYRNFNYI-YAQTIN-- 576
Db 654 -GWRYGNNSDVYVKGKPKTSRNOBFGLGKDFGNTBISHSNAYRNLAFAPELSKN 712
Qy 577 -DGRPKSIEDDSEMKLYRN--OSGADFYGAEGEYF----- 611
Db 713 GTTGKNGYHNAQAKLVGNITAIQDENGMLKRLPYGWYATFAYNRVKVKKOQKINAGL 772
Qy 612 -----KTPRYRIGVSGD-----YVRGLKMLPSLPQREDAYGNRPFI 649
Db 773 ASVSYLFDALQPS-RYIILGLYDHPSTNGINTWFTQSKAKSQNELIGQR-ALGN- 826
Qy 650 AQDQONAPRYPARLGFHLKASLTDRIDANLDYRVFAQNKLYRVEYTRPGHMLNLGAN 709
Db 827 --NSRVKSTRKLTAWHI-----LDVSGYYMANK-----NIMRLGIY 863
Qy 710 YRNTRYGEMWVVK-ADNLLN--QSVYHSSFLSDTPQMGSRFTGCVNVKP 758
Db 864 NLFNRYVTWEAVRQTAQGAHVQNGVSYRYAAS-----GRNYTLTLEMKF 911

RESULT 19
US-10-240-218-6
; Sequence 6, Application US/10240218
; Publication No. US20030186848A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: Goringe, Andrew Richard
; APPLICANT: Hudson, Michael John
; APPLICANT: Matheson, Mary Anne
; APPLICANT: Robinson, Andrew
; APPLICANT: West, David McKay
; TITLE OF INVENTION: Recombinant iron uptake proteins
; FILE REFERENCE: GMS/22057
; CURRENT FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: US/10/240,218
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-240-218-6

Query Match 3.8%, Score 151.5; DB 12; Length 915;
Best Local Similarity 19.0%; Pred. No. 66-05;
Matches 191; Conservative 131; Mismatches 345; Indels 339; Gaps 50;
Qy 1 MAQTLKPIVLSILINTPLLAQHETE---QSVGLETVTVVVGKS---RPRATSGLLH 52
Db 1 MQQHLFRFNILCISLMTALPAYAENVQAGQAEKQDITQVKAQKOTRDNVEVTGL-- 58
Qy 53 TSTASDKIL-SGDTL-RQKAVNLGDALDVPVGHASQYGGASAPV-IRGQTGRRIKV-- 107
Db 59 -----GKLVKSSDLSKEQVNLNDRTPGIAVAVQGRGASGYSIRGMKNRVSLTV 113
Qy 108 -----LNHHGETGDMADFPDHAIMVDTALSOQVEILLRGPVTLLYSSG 150
Db 114 DVSVQIQSYTAQALGGTRTAGSGGAINETIENV-----XAVEISKGSNVEQSG 165
Qy 151 NVAGLV----DVADGKIPEK-----MPENGVSGLGLRLSGNLEKLTSGGINIGLKNF 201
Db 166 ALAGSVAFQTKTADDDVIGEGRQWGIQSKTAYSGK-----NRGLTQSIALAG-RIGGAEL 219

Qy 202 VLHTEGLYRKSGDY-----AVPRYRMLKELPDS----- 229
Db 220 LIHTG--RRAGEIRAHEDAGRGVGSFNLVPEVDESSNAYFIVKEECQNGSYETCKANP 276
Qy 230 -----PRRF-----ANGQRAVL-----GWR-----KRPY-----RR 251
Db 277 KDVGVGDERQTVSTRDTGPNRFLADPLSYESRSWLFPRGFRFENKRHYIGGILEHTQQ 336
Qy 252 TYSDR-----RDYQ-LP-----ASHHEYDDCHAD----- 275
Db 337 TFDTRDMTPAFUTKAVFDANKQAGSLPQNGKYAGNHKHYIGGLFTNGENGALVGAEGYTG 396
Qy 276 IIVOKSLINKRYLQLYPHLLTEED-----VDYDNPGLSCGFH-----DDDDAHAAH 320
Db 397 VFDETHTRSYGLEVYVYTNADKDTWADYARLSYDQGVGLDNHFQOHTCSADGSKYCR 456
Qy 321 AHNGKQWIDLNRKY-----ELRAEWKQPP-----GFEALRVHLNENDYH 361
Db 457 PSADKPFYYSYKSRVYIGSHRLLQAFKKSFDOTAKIRHNLNVNGLGDFRFGNLRHQDY 516
Qy 362 HDEKAGDAVENFFNNTQNAIRLHQ-----PIGRKSGWGVQYLGQ-----KSSAL 409
Db 517 Y-CHANFAYSSNTFPQNGKKISPNGSETSPYVWTIGR-----GNVVTGQICRGNNTYT 570
Qy 410 SATSEAVKQMLLDNKVQHYSPFGVQANWNTFLEGVURVEKQASIRYDKALIDRENY 469
Db 571 DCTPRSI-----NGKSYA-----AVRDNVRLG-----RWADVAGAGLYD----- 605
Qy 470 YKQPLDPLGA-HRQTARSPALSGNMYFTPOHKLST--ASHQERLPSTOELVYAHGKHVAT 526
Db 606 YRSTHSDGVSCTGTRTLTSWAGIVLKPDTMDLTYRTSTGFLPSPFAEMYGWAGVQS 665
Qy 527 NTFEVGNKHLNERSNNIELALGYGDRQWNLALYRNFNYI-----YAQTLDG-----R 579
Db 666 KAVKI-----DPEKSFNKEAGIVFKGDPGNLEASWFNFNAYRDLIVRGYEAQIKDGKEAK 720
Qy 580 GPKSIEDDSEMKLYRVNOSG-----ADFYGAE 607
Db 721 GDPALNAQARITGINILKIDWNGWDLKPGWYSTFAYNRVRVDIKKRADRTDIOS 780
Qy 608 EYFKPTP-RYRIGVSGDYVRG-----RLKNLPSLPQREDAYGNRPPIAODQN 655
Db 781 HLFDAIQPSRYVWGLGYDQPEGKMGVNGMLTYSKAKEITEL-----LGSRALINGNSRN 834
Qy 656 APRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLYRVEYTRPGHMLNLGANYRNT 715
Db 835 TKATARTRPWYI-----VDVS-GYITV-----KKHFTLRAGVYNNLLNYR 873
Qy 716 YGEW-NWYVKADNLLNQ--SVYHSSFLSDTPQMGSRFTGCVNVKF 758
Db 874 YVTWENVRQTAGGAVNQHKVGVYVRYAAP-----GRNYTFSLEMKF 915

RESULT 20
US-09-332-226-2
; Sequence 2, Application US/09332226
; Patent No. US2002025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhoea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/332,226
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/363,124
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/973,336
 FILING DATE: 05-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/572,187
 FILING DATE: 23-AUG-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gallagher, Thomas C.
 REGISTRATION NUMBER: 37,066
 REFERENCE/DOCKET NUMBER: SPA-1-PDC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 645-1405
 TELEFAX: (212) 645-2054
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 915 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-332-226-2

Query Match 3.7%; Score 150.5; DB 9; Length 915;

Best Local Similarity 18.4%; Pred. No. 7.5e-05;

Matches 186; Conservative 132; Mismatches 344; Indels 349; Gaps 48;

QY 1 MAQTLKPIVLSILLINTEPLAQAHETE---QSVGLETVTVVGS---RPRATSGLLH 52
 DB 1 MQQHLFRNLICLSLMTALPAYAENVQQAQEKQDITQVAKKOKTRDNEVTGL-- 58
 QY 53 TSTASDKII--SGDTL--RQKAVNLGDALDGVPGIHASQYGGASAPV--IRGQTRIRKV-- 107
 DB 59 -----GKLVKTADTLSEKQVLDRLTRYDPGIAVVEQGRGASSGYSGIRGMDKNRVSITV 113
 QY 108 -----LNHHGETGDMADFPDHAIMVDLTALSOQVEILRGPTVLLYSSG 150
 DB 114 DGLAQIQSYTAQAALGGTGTAGSGAINEIEYENV-----KAVEISKNSVVEQSG 165
 QY 151 NVAGLV-----DVADGKIPEKMPENGVSGLRLSGNLEKLTSGGINIGLKNFVL--- 203
 DB 166 ALAGSVAFQTKTADVI-----GEGRWGIQ-----SKTAYSGKNRGLTQSIALAGR 212
 QY 204 --HTEGLYKSGDYA-----VPRYENLKRLPDSEPR-----RFANGQH--- 238
 DB 213 IGGAEALLIRTGTHAGEIRAHEAAGRGVQSPNRLAPVDGSGKAYAFIVEECKNGGHEK 272
 QY 239 -----RAVLG-----W-----RKRFF-----R 250
 DB 273 KANPKKVVGEDKQTVSTRTDTGPNRFLADPLSVESRWLFRPGFRENKRVIGGILS 332
 QY 251 RTYS--DRDQ-----YGLPAHSHEVD-----D 271
 DB 333 RTQQTFFTRDMTPAFUTKAVFDANQKQAGSLRGKYGAGNHKYGGLFTSGENNAAPVAG 392
 QY 272 CHADIIWOKSLINKRYLQLYPHLLATEE-----VDYDNPGLSCGFH-----DDDD 316
 DB 393 YTGCVFYDETHYSKYGLVYVYTNADKTWADYARLSYDRQAGLIDNFFOOTHCSAGSD 452
 QY 317 AHAAHNGKPMIDLNKRY-----ELRAEWKQFPF-----GFEALRVHLNR 357
 DB 453 KYCRPSADKPSYKSDRVYVIGESHKLQAFAFKKSFDTAKIRHNLVNLGYDRGSLNRH 512
 QY 358 NDYHDERKAGDAVENFNQNTQNAIRIELRHOPIGRLKSGWGVLYGQKSSALSATSVAK 417
 DB 513 QDYIY-----QANRAYSLTKPQNNKTKSPNGREKPYVW-----SIGRGNVTR 559

QY 418 QPMILLDN-----KVQHYSPFGVQQAQNDNFTLSEGGVVRVEKQKASIRYDKALIDREN 468
 DB 560 QICLFGNNTYDCTPRSGKSYAAVR---DNVRLG---RWADVAGAGLYD----- 605
 QY 469 YKQPLPDLGA--HQQTARSPALSGNWTFTPOHKLST--ASHQERLPSTOELVYAHGKHVA 525
 DB 606 -YRSTHSDGVSSTGTHRTLSWAGIVLKPADWLDLTYRTSTGFRPLSPFAEMYG----- 658
 QY 526 TMTFEVGNK-----HLNKERSNNIELALGYEGDRNQYNLALYRNRFQNYI---YAQTLDNG 578
 DB 659 ---WRSQDKIKAVKIDPEKSPKEAGIVFKGDFGNLEASWENNAYRDLIVRGVEAQIKDG 715
 QY 579 R-----GPKSIEDDSEM-----KL-----VRYNQSADP 602
 DB 716 KEQVKNPAYLNAQASARITGINILGKIDWNGVDKLEGEWYSTPAYNRVRDRIKKADR 775
 QY 603 YGAEGEIY--FKPT-----PRYRIGVSGDYVRGRLNKLPFLSPFGREDAYQNRPFIA 650
 DB 776 TDIQSHLFDALQPSRYVVGSGYDQPEGKQWGVNGMLTVSKAKEITEL-----LGSRALLN 829
 QY 651 QDDQNAQPRVPAARLGFHLKASLTDRIANLDYVYFAQNKARLYETRTPGHHMLNLGANY 710
 DB 830 GNSRNTKATARTRPWYI-----VDVS--GYITV-----KHFTLRAGVYN 868
 QY 711 RNRTRYGEW--NMYVKADNLLNQ--SVTAHSSFLSDTPQMGSRSTGQVNVK 758
 DB 869 LLNHYVTVWENVROTAAGAVNQHKVGVNYRYAAP-----GRNYTFSLEMKF 915

RESULT 21

US-10-043-344-96

; Sequence 96, Application US/10043344

; Publication No. US2003008086A1

; GENERAL INFORMATION:

; APPLICANT: Loomore, Sheena M.

; APPLICANT: Harkness, Robin E.

; APPLICANT: Schryvers, Anthony B.

; APPLICANT: Chong, Pele

; APPLICANT: Gray-Owen, Scott

; APPLICANT: Murdin, Andrew D.

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES

; FILE REFERENCE: 1038-1221 MIS

; CURRENT APPLICATION NUMBER: US/10/043,344

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: 08/649,518

; PRIOR FILING DATE: 1996-05-17

; NUMBER OF SEQ ID NOS: 160

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 96

; LENGTH: 915

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-10-043-344-96

Query Match 3.7%; Score 150.5; DB 15; Length 915;

Best Local Similarity 18.4%; Pred. No. 7.5e-05;

Matches 186; Conservative 132; Mismatches 344; Indels 349; Gaps 48;

QY 1 MAQTLKPIVLSILLINTEPLAQAHETE---QSVGLETVTVVGS---RPRATSGLLH 52
 DB 1 MQQHLFRNLICLSLMTALPAYAENVQQAQEKQDITQVAKKOKTRDNEVTGL-- 58
 QY 53 TSTASDKII--SGDTL--RQKAVNLGDALDGVPGIHASQYGGASAPV--IRGQTRIRKV-- 107
 DB 59 -----GKLVKTADTLSEKQVLDRLTRYDPGIAVVEQGRGASSGYSGIRGMDKNRVSITV 113
 QY 108 -----LNHHGETGDMADFPDHAIMVDLTALSOQVEILRGPTVLLYSSG 150
 DB 114 DGLAQIQSYTAQAALGGTGTAGSGAINEIEYENV-----KAVEISKNSVVEQSG 165
 QY 151 NVAGLV-----DVADGKIPEKMPENGVSGLRLSGNLEKLTSGGINIGLKNFVL--- 203


```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10132
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10132

Query Match      3.5%; Score 141.5; DB 9; Length 729;
Best Local Similarity 20.0%; Pred. No. 0.00038;
Matches 151; Conservative 93; Mismatches 275; Indels 235; Gaps 35;

QY 3 QTTLKPIVLVS--ILLINPTLLA-QAHETQSV---GLETVVVVGKSRPRATS-----GLLHT 53
Db 13 QAITKPSLLAGGIALALLPSAFAAPATEETVIVEGSATAPDDGENDSVTSTAGTKMQ 72
QY 54 STASD-----KTIISGDTLR-QXAVNLGDALDGPCHASQYGGASAPVIRGQTRRIKV 107
Db 73 MTQRIIPQSVTVISQORMEDQQLQTLGEMENTLGISKSQADSRLAYISRG-----PQI 127
QY 108 LNHGGETGMADSPDH-----AIMVDTALSOQVEILRGVTVLLYSSGNVAGLVDA 159
Db 128 DNY-----MVDGIPTYPEBSRWNLGDALSMALFERVEVRGATGLMTGTGNSPAAINNV 181
QY 160 DGIKEKMPENGVSGLGLRLSSNLEK-----LTSGGINIGLKNFVHLHTEGLVRK 211
Db 182 RKHATREFKGVDSAYG-----SWNKERVADLQSLPTEHG-----KIRARIVGGY-Q 229
QY 212 SGDYAVPRYRLKRLPDSPPRFANGCHRAVLGWRKRF-----YRTYSDRDQVGLP--- 263
Db 230 NNDMSLDRY-----NSEKTFPGIVDADLGLTLTLAGYERYQIDVNSPTWGLPRWN 282
QY 264 --AHSHEYDDCHA--DIIWQSKLNRKYLQLYPHL-----LTBEDVDYDNPGLSG 309
Db 283 TPGSSNSYDRASTAPDWAYNDKEINKVFWTLKQOFADTWQATLNAHSEVEFFDSKMIV 342
QY 310 GFHDDDDAHAAHNGKPTWIDLRKRYELRAEWKQFPFGFEALRVLNRNDYHHDEKAGDA 369
Db 343 -----DAVVKADGN-----LVGPYSNYPGPGF-----IVGGTWNSSGKRKYDA 381
QY 370 VENPF-----NMOTQARIELRHQIGRL-----KSGMVG 399
Db 382 LDLEADGYSYELFGROHNLMPFGYSYKQNNRYFSSWANTPPDDEIGSFYFNFGNFPQTDWSP 441
QY 400 QYLGO-----KXSALSATSEAVKQPMLDNKNVQHSFFGVEQANNDNFTLEGVVRVEKOK 454
Db 442 QSLAQDDTHMKSLYATATVTLADPLHL-----ILGARYTNW----- 478
QY 455 ASTRYDKALIDRENYKQPLDGLGAHQRTARFALSNGWYFTPOHKLSTLASHOERLPST 514
Db 479 ---RVDTLAYSMEKNTHTPYAGL-----VFDINDNW-----STYASYTSIFQFPQ 519
QY 515 QELYAHGKHVATNTEFVGNKHLNKRSSNNIELAL--GYEGDRWQYNLALRYRNFNGYIYA 572
Db 520 NORDSSGKYLAPIT-----GNNYELGLKSDMWNRLTTTLAIFR----- 558

```

```

QY 573 QTINDGRGPKSIEDSEMKLVRVNSGADFYGAEGEIVYKPTPRYRIGVSGDYVRGLKN 632
Db 559 -----TEQDNVAQ-----STGTPIPSNGGETAYK-----AVDGTVSGKGEFE 595
QY 633 LPS--LPGREDAYGNRPFTIAQDDQ-----NAPR 658
Db 596 LNCALTNDNQLTFTGATRYIAEDNEGNAVNPMLPR 629

RESULT 25
US-09-881-752A-334
; Sequence 334, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 791
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-881-752A-334

```

```

Query Match      3.3%; Score 135; DB 10; Length 791;
Best Local Similarity 19.4%; Pred. No. 0.0018;
Matches 170; Conservative 112; Mismatches 352; Indels 242; Gaps 46;

QY 19 PLLAAQHETQSVGLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLRKAVNLGDALD 78
Db 16 PLLAETCYSEBT-LNKVTTQAKR-----IFTYNEFKVTSKELDQOSNEVDLFR 66
QY 79 GVPCHASQYGGGA---SAPVIRGQTRRIKV-----LNHGETGMADSPDH 124
Db 67 TNPEDVNV---GGGSVMGQKIYVRGVEDRLRLRVTDGAAQNGNIYHHQGN----- 113
QY 125 AIMVDTALSOQVEILRGVTVLLYSSGNVAGLVDAVDGKIPEKMPEN---GVSGELGLRLS 181
Db 114 --VIDPMLKSVEVTGKAANASAGPAGIAGVKNMETYGAADPIRGKNYASGAVSYTN 171
QY 182 SGNLEKLTSGGINIGLG-KNFVLHTBGLYRKSGDYAVRYRN---LKRLPDSPPRFANGQ 237
Db 172 FGORETFRSAQNAHFIIAYYTHQIFVYRSGATAMKLNFNPTQADKEPGTP---SEQ 227
QY 238 HRAVLGWRKRYAR-----TYSDRRQYGLPAHSHEVD---DCHADIIWQSKLNRKYLQ 289
Db 228 NNALIRKNGYLSRDRTLTFSWNMTNRNATRLSRNAIGLAIYPCAPFSPSSQGCNVLD 287
QY 290 LYPHLLTEEDVDYDNPGLSGCFHDDDAH-----AHAIN--GKFWID-----LRNKR 334
Db 288 SFTRYM-----YHSINSANNLSLQYKREAGNSFGDPRLDFTLTSIRNAQ 332
QY 335 YE-----LRAEW-KQFPFGFEALRVLNRNDYHHDEKAGDAVENFNNQT 378
Db 333 FDPLEDPNGYVAKFPTSLASAMEKENYPCVEGAYCTPFSFSDV---DKPSSQPNLFINNT 389
QY 379 QNARIELRHQPIGLKGS---WGVQY-----LGOKSSALS 410
Db 390 -GLNLKVAH-VIDEATDSLFEYFNQYONLSVFPDARIPKSELYRPNQVYTDKQKQIACS 447
QY 411 ATSEAVKQPMLLD-NKVQHSYFFGVEQANWDN---FTLEGVVRVEKQASIRYDKALIDR 466

```


Db 448 LVNNNEDEFLCORGKANGNIYGYQANYSPHKIITFGAGVRWD---AYTLYDK---DW 501
 Qy 467 ENYKQPLDGLGAHRTASRALSGNWFYTPQKLSLTASHOERLPSTQELYAHGKHVAT 526
 Db 502 NHRYTQ---GFSALALVLSPIEPLSLKITYSQ---VTRGVWPGDGVYMRQ 546
 Qy 527 NTEVEGNKHLNERSNNIELALGYEGDRWOYNLALYRNFRNGYI---YACTLNDGRGPKSI 584
 Db 547 NDLRYA-KNIKPEVGSNAEFNIDYSSQYFSGRAAFYQALDNFISQYAGNL----- 596
 Qy 585 EDSEMKLVYNGSGADFFVGA---GRIYFKPTPRYRIGVSGDY---VRGRLKNLPSLPGR 640
 Db 597 ---IVNLQAIRIYGYEGVGFPRYKGV-SLNVGVSRTWPTTRGYL-----MA 640
 Qy 641 DAY-----GNRPPIAQQDQNAVRPAARLGHFLKASLTDRIDANLDY---YRVFAQNKIA 692
 Db 641 DSYELAASGTGN-VFIILKDYTIPTK-----GINL---AWLSRFVTGLDYCGFDIYLYDGT 692
 Qy 693 RYETRTP-----GHEMLNIGANVRRNTRYGEMNWTYVK-----ADNLL 729
 Db 693 AERKPTITDLAKGS---QLGVHMKPGYGVSNFVNNWSPKTSKRWKGLLSAVENNVF 749
 Qy 730 NQ-----SVYHSSFLSDTPQGRSFT-CGVNVKP 758
 Db 750 NKFYVDQTSYVWSPMDPGTDAVKRAIAEPGFNARF 785

RESULT 26

US-09-815-242-5106
 ; Sequence 5106, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5106
 ; LENGTH: 815
 ; TYPE: PR1
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-815-242-5106

Query Match 3.1%; Score 127; DB 9; Length 815;
 Best Local Similarity 17.4%; Pred. No. 0.011;
 Matches 135; Conservative 95; Mismatches 242; Indels 302; Gaps 36;
 Qy 21 LAQAHETEQSGVLETWT-----VVGKSRPRATSGLLHTST-----ASDKIISGDT 65

Db 117 VAEAAUSSVDLGMTITSNQLOGTITDSQYTPGTIATATRLVLTPTRETPQITVVTQON 176
 Qy 66 LROKAV-NLGDALDGVPGIHASQYGGASAPVIRQOTGRRIKVLNHHG---ETGDMADSP 122
 Db 177 MDDFGLNIDDDVRHTFGITVSAYDTRNNYVARGFS---INNFOYDGIPTARNVGYSA 233
 Qy 123 DHAIMVDTALSOQVEILRGVTVLLYSSGNVAGLVADVADCKIPEKMPENGVSGLGLRLSS 182
 Db 234 GNTLS-DMAIYDRVEVLKGAATGLTGAAGSLGATINLI-----RKPTEFKGH--VELCA 285
 Qy 183 GNLEKLTSGGINICLQKGNFVLHTGELYRKSGDYAVPRYRNKLKELPDSPRRFANGHAYL 242
 Db 286 GSWDN-----YRSELDVSGP-----LTES-----GNVRGRAVA 313
 Qy 243 GWRKFRYRTYSRRDQ---YGLPAHSHVEDDCHADIIWQSLINKLEYLOLYPHLLTEEDV 300
 Db 314 AYQDKHSMFMDHYERKTSVYVYGI-----LEFD-----LNPTMTVTGA 350
 Qy 301 DY-DNPGLSGCFHDDDDAHAAHNGKPIDLRNRYELRAEWKQPPPGFEALRVHLNRND 359
 Db 351 DYQNDPKGSG-----MSGSPFLFDS---QGNRND 377
 Qy 360 YHDEKAGDAVENFNNQTONARIELRHQPIGRKSGWVOYLQOKSSALSATSSEAVKOP 419
 Db 378 -----VSRSFNN----- 384
 Qy 420 MLLDNKVQHYSGVEQANWDFTEGGVREVKQKASIRYDKALIDRE-NYKQPLDGLG 478
 Db 385 -----GAKWSSWEQYTRTVFANLEHNFANGVGVQVLDHKNINGYHAPLG--- 428
 Qy 479 AHROTARSPALSGNNYTPQHKLSLTASHQ---RELSTQELYAHGKHVATNTE-VGNKH 535
 Db 429 -----AIMGDWP-APDNSAKIVAOKYTGTGKNSLDIY-----LTGPPQFGLGREH 472
 Qy 536 LNKERSNNIELALGYEGDRWQYNLALYRNFRNGYVIAQTLDNCRGPKSIEDDSEMKLVRY 595
 Db 473 -----ELVVGTSAS-----FSHW-----EGKSYWNLRY 496
 Qy 596 NOSGADFYGAEGEYFRTPRYRIGVSGDYVRGRLKNLPSLPORDAYGNRPFIADQDN 655
 Db 497 DNTTDDFINWDGI-GKPD-----WGTFESQYIDDKTROLGS-----YMTARENVTTDD-- 542
 Qy 656 APRVPAARLGHFLKASLTDRIDANLDYRVFAQNKIARVETRTPGHMLNMGANRYENR 715
 Db 543 -----LNLFLGGRVVD-----YRVTLNLP-----TIRSGRPIPVGAYVLDNDT 582
 Qy 716 YGEW-----NMVVKADNLLNQSVAHSSFLSDTPQMGSRFTGGVNVKP 758
 Db 583 YSVVASYTDIEWPDQSWYRDSNNKLE-----PDEGQNYEIGIKGEY 624

RESULT 27

US-09-881-752A-362
 ; Sequence 362, Application US/09881752A
 ; Patent No. US20020115078A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kleanthous, Harold
 ; APPLICANT: Al-Garawi, Amal
 ; APPLICANT: Miller, Charles
 ; APPLICANT: Tomb, Jean-Francois
 ; APPLICANT: Oomen, Raymond P.
 ; TITLE OF INVENTION: Identification of Polynucleotides
 ; TITLE OF INVENTION: Encoding No. US20020115078A1
 ; FILE REFERENCE: 06132/041002
 ; CURRENT APPLICATION NUMBER: US/09/881,752A
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: US 08/833,457
 ; PRIOR FILING DATE: 1997-04-01
 ; NUMBER OF SEQ ID NOS: 370
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 362

Query Match	3.0%	Score 122.5	DB 12	Length 767
Best Local Similarity	18.0%	Pred. No. 0.027		
Matches 161	Conservative 117	Mismatches 337	Indels 279	Gaps 42
Qy	14	LLINTPLLAQAHETEQSGLVETTVVG----	KSRPRA-TSGLLHTSTASDKIISGDTLURQ	68
Db	4	ILVSLAVLSHSAHVKTHNLERVEASGVANDKEAPLSWESKEVRNMGSRVVISNKQLTK	63	
Qy	69	KA-VNLGDALDGVPGIHAHQ-----	YGGGASAPVIRGQTGRRIKVLNHHGE	113
Db	64	SANQSIIEALQNVPGVHIRNATGIGAVPSPFVRGFGGSS-----	CHSNMTAMLVN-----	114
Qy	114	TGDMADPSP--DHAIMVDTALSOQVE--	ILRGPTLLYSSGNVAGLVADVADGKIPKWP	168
Db	115	-GIPYIVAPYVDISIPFPYTFQSVDRI	SVTKGESVRYGNPVGVGINVITKGIPTKW-	172
Qy	169	ENGVSSELGL--RLSSGNLEKLTSGGINTLGKNFV----	LHTEGLYRKSGDYAVPRYN	222
Db	173	ESQSERATFWKSGENGFFNQSNKLDKSLANMLFDYTLRTGGMMNK--	HFGIQAQAN	230
Qy	223	LKRLPDSRRFANGQRAVLGWRKFRYRTYSRRDQYGLPAHSHEYDDCHADIIWQKSL	282	
Db	231	-----WLKQO-----GFR-----	YNSPTNIQNY--MLDSYIQIND	258
Qy	283	INK--RYLQLYPHLLTBEDVDYDNPGLSCGFDHDDAHAAHNGRPWIDLNRKRYELRAE	340	
Db	259	SNKITAFFQYINFMAD-----PGSL-----		279
Qy	341	WKQPPGFEALRVHLNENDYHHDEKAGDA-----	VENEFNNQTNQARTELRHQPIGRLK	394
Db	280	-----GLEAYNQRFNNPNNKSGRAKKGAVIQNFQD-----	TDKIG	320
Qy	395	GSWGVOYLQO-----KXSALSATSEAVKQPMLLDNKVQHSYSPFG-----		433
Db	321	GDFTSYXGHDMSRDFQDFSNFLNVTNPKLGPVYTDQNVPGFFIDHLRLRYIMNAPEPN	380	
Qy	434	---VEQAWNDNFTLEGGRVVEKQKASIRYDKALIDRENYKQ--	PLPDLGARHQTARSA	488
Db	381	LNLVNTNKKVQTFNVGMRFWMTQMYPRLOQSTCEKTDIFNGVCRMPPLVSKFSSNNQN	440	
Qy	489	LSGN-----WYFTPOHKLSTASHOERLPSTQELYA--	HGKHVATNTFEVGNKHLNKRSN	542
Db	441	LFNNYTAVMWL--SDKIETPLSKLIVTPGLRYTFLNYYNKEPKHDFSVN--	ITKKRQN	495
Qy	543	NIELAL--GYE-----	GDWRQWY-NILA	560
Db	496	ENSPALNIGYKPMENWIWYANRRSFPTPOHTMLGTRINYQINFEIIEVGQRYISKMLL	555	
Qy	561	LYRNR-----FGNYIYAQTLDNGRGPKSIEDSEMKLVRYNQSGADFYGAEGGIYFKPTPR	616	
Db	556	SFTNYFVIFAKRYYA--GGYSQPQFINARSQ-----	GVELELYVAPIRG	597
Qy	617	YRIGVSGYVVEGRU--KULPSLPGREDAYGNRPPIAODQONAPRVPAARLGPHLKASLTDR	675	
Db	598	LOFHVAYTYIDARITSNADDIAYFTGIVNKKPFDIK--GRKLPYVSGNPFIDFMVYTKHT	656	
Qy	676	IDANLDYRVYRFAQNKL--ARYET-----RTPGHMLNLGAN-----	YRRNTRY	716

D0	441	LFNFNTIAWVL---	SUKIELFDSKLVITPGUKYTFPLWINNKPCKHDFSVMN---	LTKKQCN	499	
Qy	543	NTELAL---GYE---	-----	GDRWOY-NLA	560	
Db	496	EWSPALNIGYKPMENWIWYANRSEFIPPOHTWLGITRTNYNQIFNWEIVGQYSYKNLL	555			
Qy	561	LYRNR-----	FNQYIYAOTLNDGRGPKSIEDSEMKULVRYNQSGADPYGABEIIYFKPPTPR	616		
Db	556	SENTNYVIFAKRYYA---	GGYSQPQINARSQ-----	GVELEIYYAPIRG	597	
Qy	617	YRIGVSGDYVRGRL-KNLPSLPGREDAYGNRPETIAODQNAQPVPAARLGHKLKASLTDR	675			
Db	598	LOFHVAYTVIDARITSNADDIAYFTGIVNKKPDKR-GRRLPVSPNQFIFDMYTYKHT	656			
Qy	676	IDANLDYVRVFAQNKL---	ARYET-----	RTPGHMLNLCAN-----	YRRNTRY	716

Db 657 TFGISSYFYGRAYSSMLNQAQSQTVCLPLNPEYTG---LEYCNSVGLLPLFLVFLNVQV 713
 Qy 717 GEMWNY-----VKADNLLNQSVYAHSSFLSDT---PQMGSRSTGGVNVKFP 758
 Db 714 SSVLMQSRHKITGSLQINNLFNNKYFYRIGTSPTRGPAPGRSITAYLVNEF 767
 RESULT 29
 US-10-180-326-1
 ; Sequence 1, Application US/10180326
 ; Publication No. US20030049661A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Seino, Susumu
 ; APPLICANT: Shibasaki, Tadao
 ; APPLICANT: Ozaki, No. US20030049661A1uaki
 ; TITLE OF INVENTION: Protein Rim2
 ; FILE REFERENCE: P21573
 ; CURRENT APPLICATION NUMBER: US/10/180,326
 ; CURRENT FILING DATE: 2002-06-27
 ; PRIOR APPLICATION NUMBER: JP 288372/99
 ; PRIOR FILING DATE: 1999-10-08<160> 5
 ; NUMBER OF SEQ ID NOS:
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 1590
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-180-326-1

Query Match 3.0%; Score 122.5; DB 15; Length 1590;
 Best Local Similarity 18.6%; Pred. No. 0.09;
 Matches 139; Conservative 102; Mismatches 286; Indels 221; Gaps 35;
 Qy 63 GDTLRKAVNLGDALDVGPGI-HASQYGGGASAPVIRGQTGRRIKVLNHHGETGDMADFS 121
 Db 116 GEESQOQEQKGA--PTGIGCHTKFADGCHNCSCYQI-----KFCARCGGRVSLR 166
 Qy 122 PDHAIMVDTALSQVEILRGPVLLYSSGNVAGLVADVADGKIPEKMPENGVSSELGLRLS 181
 Db 167 SNKVMVYCNLCRQKEILTSGAWFNSG--SNTLQPDQKQVPR-----GLRNE 213
 Qy 182 SGNEKLTSGGINIGLGNFVLTGELYRKS-GDYAVP-----RYRNLKRLPDSRRFAN 235
 Db 214 EAPQEKAK-----LHQPQFQAGDLSVPVKEGGRACLTR-QDTIKNGSG 260
 Qy 236 GQRAVLGM--RKR-----YRTY--SDRRQYQ-----LPASHVYDDCHADI 276
 Db 261 VKHQIADMPDRKSPSVSRDQNRREYQSEEREDYSQYVPSDGTWPRSPSYAD----- 315
 Qy 277 IWQKSLINKRYLQLYPHLLTEEDVDYDNPQ--LSCGFHDDDDAHAAHNGKPWID----- 329
 Db 316 -----RRSQREPOF-----YEPGHNLN---YFDSNRCHRHRSKEYIYVDDDEVES 356
 Qy 330 -----LRNKRYELR-----AEWK-QPPGPEALRVHLNRNDYHHDEKAGDAVENFF 374
 Db 357 RDEYERQRREYQARYSPNRLARYPKQPYEQVBEQRIHAESRARRHRSDDV--SLA 414
 Qy 375 NNQTONARIEL-----RHQPIGLKSGMVGQVILGOKSSALSATSAAVKQ----- 418
 Db 415 NAELEDSSILRLMDRSPRQBSVERRAWE-----NORSYSWETREAAQSSYSPORTSN 470
 Qy 419 -----PMLLDNKVQHSYFFGVEQANWDMFTLEGGRVVEKOKASIRYDKALIDRENY 469
 Db 471 HSPPTPRSPITLDRPMRRADSLRKQHLLDPSSAVRKTREKMETMLRNDLSLSQDS 530
 Qy 470 YKQPLPDLGAHQRTARSFALSGNWFYFPQHLKSLTASHQERLSTQELVYAHGHV---AT 526
 Db 531 VRPPPPRPHKSKGKGK-----MRQVSLSSS--EELASTPB-YTSCDDVELESE 576
 Qy 527 NTFVGNKHLNKRNSNIELALGYEGDRWQYNLALYRNFQNYIYAQTLNDGRGPKSIED 586
 Db 577 SVSEKGDQSKGRKTSE-----QGVLSDSNTSRSEKQKRMYY-----GGHSL 620

Qy 587 DSEMKLVRYNOSGAD-----FYGABGEIYFKPTPRYRIGVSGDYVRGRL---XN 632
 Db 621 DLEWSEPOIKDSGVDTCSSTLTNEEHSRDKHPVTWQPSK-----DGRLLIGRLINKR 674
 Qy 633 LPSLPGRDAYGNRPFFIAQDDQNAFPRVPAARLGFHLKASLTDRIDANLDYYRVFAQNKLA 692
 Db 675 L-----KDSVPRDSGAMGLGLKVVGG-----KMTESGRLC 704
 Qy 693 RYETRTPGHEMLNLGANYRNRTRYGEWN 720
 Db 705 AFITKVKKGLADTVGHLRPGDEVLEWN 732
 RESULT 30
 US-10-238-075-1277
 ; Sequence 1277, Application US/10238075
 ; Publication No. US20030148124A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol
 ; FILE OF INVENTION: E.coli, and biological uses of these polynucleotides and of th
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; PRIOR FILING DATE: 2000-03-10
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1277
 ; LENGTH: 753
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-10-238-075-1277

Query Match 3.0%; Score 120.5; DB 12; Length 753;
 Best Local Similarity 20.8%; Pred. No. 0.041;
 Matches 146; Conservative 90; Mismatches 283; Indels 183; Gaps 34;
 Qy 30 SVGLETVTVVGKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDVGPGIHAS-QY 88
 Db 44 SVGKTTSEALEKTGATS-----SRTTDKNLQ-----SLDATVRSMPGTVTQIDP 89
 Qy 89 GCGASAPVIRGQTG-RRIKVL-----NHHGETGDMADFSDPHAIMVDTA 131
 Db 90 GCGAISVNRIGNSGFRVNTWVDGITSQFYGTSTGTTTHGSTNNMA-----GVLIDPN 143
 Qy 132 LSQOQVEILRGP-----VTLLYSGNV--AGLVADVADGKIPEKMPENGVSSELGLRLSSG 183
 Db 144 LLVAVDVTRGDSGSGSEGINALAGSANMRTIGVDDVI-----FNGNTYGLRSRFSVG 194
 Qy 184 NLEKLTSGGINIGLGNFVLTGELYRKS-GDYAVP-----LYRKSQDYAVPRYRNLKRLPDSRRFANGQHR 239
 Db 195 SNGLGRSGMIALG-CKSDAFTDTGSGYMAAVSGSSVYSNFSN-----GSGINS 242
 Qy 240 AVLGRKRFYRRTYS-----DRRDQYGLPAHSHEYDDCHADI IWQKSLINKRYLQ 290
 Db 243 KEFGYDKYMKQNPQSOLYKMDIRPDEFNSFELSARTYENKTRDI-----TSDDYIK 296
 Qy 291 YPHLLTEEDVDYD-NPGLSCGFHDDDDAHAAH-----AHNGKPWIDLRN-KRYELRAEWK 342
 Db 297 YHTTFFSELIDFNVTASTSRGNQKRYDRSLTYFKTSAQNRSDALDINNTRSRFTVADNDL 356
 Qy 343 QPFPGFPEALRVHLNRNDYH--HDEKAG--DAVEN--FFNNQTONARIELRHQPIGLKSGW 397
 Db 357 EFWLGSKLMRTYDRTIHSAGDPKANOESTENNPFASGGQD--ISALYGLKVTGRIW 414
 Qy 398 GVOY-LOGKSSALSATSBAVKQPMLLDNKQVHSYFFGVEQANWDMFTLEGGRVVEKOKAS 456
 Db 415 EADFNLYVTRNRIITGYKPA-----CDSRV-----ICVPQGSYDIDDKEGGF----- 455
 Qy 457 IRYDKALIDRENYKQPLPDLGAHQRTARSFALSGNWFYFPQHLKSLTASHQERLSTOE 516
 Db 456 -----NPSVQLSAQ-----VTPWLPQFFIGYSKMRAPNIQE 486

QY 517 LYAHGHVATNTFEVGNKHLNKRNS-----NIELA-LGVEGDRMWNLYALYRNFRGNY 569
Db 487 MFFNSGGASM-----NPLKPERAETWQAGFNIDTRDLLVEQDALKALAYRSIQNY 541
QY 570 IYAAO-----TLND--GRGPKSIEDDSEMKLVRYNOSGADFYCAGBEIYFKPTPR 616
Db 542 IYSESVLVCGRKSLPEVINGWEGISDESDNMYIYVSASDVI--AKG---PELEMD 597
QY 617 YRIGVSGDYVRGLKNLPSLPGRDAYGNRPPIAODQNA 658
Db 598 YDAG-----FAFGRLSPSQOQDPTQPTSIASHTFGAGDITELPR 635

RESULT 31
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 2.9%; Score 118.5; DB 9; Length 1475;
Best Local Similarity 18.7%; Pred. No. 0.19; Indels 227; Gaps 35;
Matches 125; Conservative 85; Mismatches 230;

QY 87 QYGGGAPVIRGQTGRRIKVLN--HHGETGDMADFPDHAIMV---DTALSQQVEILRG 141
Db 894 RYDLGSKPNKYGTADLVKAIKALHSKGIKWADWVPDQMYAPPEKEVVTATRVDKYGT 953
QY 142 PV-----TLIYSSGNVAG-----LVDVADGKTPE-----KMPENGYSSELGLRL 180
Db 954 PVAGSQIKNTLYVVDGSSGKQOQAKYGGAFLELOAKYFELPARKQISTGVPMDSVKI 1013
QY 181 SSGNLEKLTSGGINIGLQKQFVLHTEGLYRKSGDYAVPRYRNL---KRLPDGPRRPAQ 237
Db 1014 QKWS-AKYFNGTNIIRGAGVVLK-----DQAINTYFNISDNKEINFLPKTLNQD 1063
QY 238 HRAVLGWRKRY-----RRTYSDRDRDQVGLPAHSHEYDDCHADIIWQKSLINKRY 287
Db 1064 SQVGFSDYDGKGVYVYSTSGYQAKNTFISEGDKW-----YFQNNNGYMTVGAOSINGVNY 1117
QY 288 -----LQLYPHLLATEEDVDYDNPGLSCGPHDDDDAHAAHNGKPMIDLANKRYE----- 336
Db 1118 YFLSNGLOLRALKKNEDGTY-----AYGND-----GRYENGYQ 1154
QY 337 -LRAEWKQFPFGFALRVHLNRND-----VHDE-----KAGDAVENP 373
Db 1155 FMSGVWR-----HFNNGEMSVGLTVIDGQVQYFDEMGMGYQAKGFVTTADGKIRY 1203

QY 374 FNNOTONARIELRHQPIGLKSGWGVQYLGOKSSALS-----ATSEAVKQPM 420
Db 1204 FDKQSGNM---YRNRFIENEKGW---LYLGEDGAAVTSQTINGOHLIFRANGVOVKGEF 1258
QY 421 LLDN--KVQHYS-----FFGVEQANWDF-----TLEG-----GVR 449
Db 1259 VTDHGHRI SYDGNSGDQIRNRFVNRNAQGWFFDNNGYAVTGTARTINGQLLYFRANGVQ 1318
QY 450 VEKQKASIRYDALKIDRENYKQPLDLCGAHPOTARSEA--LSGNWYFTPOHKLISLTASH 507
Db 1319 VKGEFVTDY-----GRISYIY-----DGNSGDOIIRNRFVRNAQGWFFDNNGYAVTGA- 1367
QY 508 QERLPSTQELIYAHGKHVATNTFEVGNKHLNKRNSNIELALGYEGDRMWNLYALYRNRF- 566
Db 1368 --RTINGQHLIYFRANGVQVKGEFVTDH-----GRISYDGNSGDQ-----IRNRFV 1412
QY 567 -----GNVIYATLNDGRGPKSIEDSEMKLVRYNOSGADFYCAGBEIYFKPTPRIGWS 622
Db 1413 RNAQGWFFD--NNG-----YAVTGARTINGQ-HLYFRAN---GVQVK 1450
QY 623 GDYVRGR 629
Db 1451 GEFVTDH 1457

RESULT 32
US-09-815-242-5247
; Sequence 5247, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5247
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5247

Query Match 2.9%; Score 117.5; DB 9; Length 422;
Best Local Similarity 18.3%; Pred. No. 0.031;
Matches 91; Conservative 69; Mismatches 149; Indels 189; Gaps 21;

QY 61 ISGDTLRQKAVNLGDALDVGFIHASGYGGASAPVIRGOT--GRIKVLNHHGETGDMA 118
Db 27 VDTDLRLKLABQVDDI-----VFISGTNGKTTTSLNIGHTLKANNIQI--HNNEGANMA 79

Qy 119 DFGPDHAIWDTALSOQVEILRGVTLTYSSGNVAGLVADGKIPKMPENGVS-----173
 Db 80 AGITSAFIMQST-----PKTKI-----AVIEIDEGSIPLVLKEVTPSMVFT 121
 Qy 174 -----GELGLRSLSSGNLEKLTSGGINIGLGNFVLHTEGLYKRGDVAVERNL 223
 Db 122 NFRDQMDRGEIDIMWNN-IAETISNKGKIKLL-----NADDFVSRLL-----164
 Qy 224 KRLPDSPPRFANGQRAVLGRKFRYRTYSDRRDQYGLPAHSHEYDDCHADIIWQKSLI 283
 Db 165 -----KIASDTIVVYGMKAHAHEPE-----QSTWN 189
 Qy 284 NKRYLQYPLHLLTEEDVDNFG-----LSCGFHDDDDAHAAHNGKRWIDLRNKRVELRAE 340
 Db 190 ESRYCPNCGKLLQVYIHYNQIGHYHCQCGKKE-----QAKYEISF 232
 Qy 341 WKOPFGFEALRVHLNRNDYHDEK-AGD-----AVEN-----372
 Db 233 DVAPF-----LHLNINDEKYMKTAGDFNAYNALTAVTLRELGLNEQAIKNGFETVT 285
 Qy 373 -----FFNQTQNAIELRHOPIG-RLKSGWGVQYLGOKSSALSATSEAVKQPMLDN 424
 Db 286 SDNGRMQYFKKERKEMINLAKNPAGMNASLSVGEQLEGEKVVIS-----LND 335
 Qy 425 KVQHSYFPGVEQANWDMFT-----LEGGRVVEKOKASIRYDKA-----LIDRENYKOP 473
 Db 336 AADGRDTSWIYDADFELKSQOIEAIIVTGTAEELQLRLKLAEEVPIIVERDIY-----391
 Qy 474 LPDLGAHRCARSFALSG 491
 Db 392 -----KATAKTMDYKG 402

RESULT 33

US-09-815-242-5456
 ; Sequence 5456, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ IDS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5456
 ; LENGTH: 978
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5456

Query Match 2.9%; Score 117; DB 9; Length 978;
 Best Local Similarity 18.9%; Pred. No. 0.14;
 Matches 162; Conservative 103; Mismatches 277; Indels 316; Gaps 44;
 Qy 19 PLLAAHETQSQVLETV-----TVVGKSRPRATSGLLHTSTASDKIISGDTLQKQAVNLGD 75
 Db 73 PKAVQAPQAPQANLETIVKEEVKERAKPQ-----VKETTPQDN--SGD--QKQVDL-- 121
 Qy 76 ALDGVPGIHASQYGGGASAPVIRGQYGRRIKVLNHHGTEGDMADFPDHAHWDTALSOQ 135
 Db 122 -----TPKMATQ-----NQAAET-----Q 135
 Qy 136 VEILRGVTLTYSSGNVAGLVADGK-IPKMPENGVSCELGLRLSSGNLEKLTSGGIN 194
 Db 136 VEVAQ-PRTVSESNPRVRSADVVKEASDAKVEGTVDVTSKVTVSSGSEIAEPQGNKVE 194
 Qy 195 IGLGNFVL-----HTEGLYKSGD-----YAVPRYRNKRLPDSPPRFANGQ 237
 Db 195 PHAQQRVVLKYKLFKEKGLHK--GDYDFFTLSNNVNTYGVSTARKVPEIKNGSVVMATGQ 252
 Qy 238 HRAVLGWRKFRYRTYSDRRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQYPLHLLTE 297
 Db 253 ---LLGNK--IRYTFDYID-YKVNVT-----DLEINLFIDPKTVQSGQQTITSLND 302
 Qy 298 ED-----VDYDNPLSCGPHFDDDDAHAAHNGKPMIDLNRKYE-----336
 Db 303 KETKNTLPIEY-NEGVS-----NSYANVNGSIEFTDKGNRFTTHVAVIKPONGHKSDS 354
 Qy 337 -----LRAEWKOP-FPGFEALRVHLNRNDYHDEKAGDAVENFFNQTQNAIE 385
 Db 355 VSIQTTLTQSGADGKAPTQVYEVLR-----DANELFQSVYANVSDSMFADVTEEMKDL 411
 Qy 386 RHQ-----PIGLKSGWGVQYLGOKSSALSATSEAVKQPMLDNKNVQHSYFGEV---435
 Db 412 KVENNGNVKLDIEKLEKSYVIHYDGE--YLSGSDQVNFRTMHFGYPEQYKYVYTHLGY 468
 Qy 436 QANWDM-----FTLEGGRVVEKOKASIRYDKALIDR-ENY 469
 Db 469 KLTWDNGLVLYSNKAKGDTNGTITSNMNTFDEEYGTGV-----ITGVQDKNLVTTVEE 534
 Qy 470 YKQPLPOLGAHR-----QTARFALSNGWYFTFQ-----HKLSTASHQE 509
 Db 525 YDSSTLIDYHTAIDGEGYVDGYETIETDSSAIDYHTAVDSEAGHGVGYTESSEE 584
 Qy 510 RLP-----STQELYAH-----GKIVAT--NTFEVGNKHL-----536
 Db 585 SNPIDFEESTHENSKHADVVVEEDTNPGGQVTTESNLVFEFDESIGIVTGAVS DHT 644
 Qy 537 ---NKE---RSNNIEL-----ALGYEGDRWQYNLAL 561
 Db 645 TVEDTKEVTTESNLIELVDELPEHGOAQGPBEITENNHHISHSGLGTE-----694
 Qy 562 YRNRFNVIYAQTLD-----GRGPKSIEDDEMKLVRNQS-----AD 601
 Db 695 --NGHGNVGVIEEIEENSHVDIJKSELGYEGQNSQNSQNSFEEDTEEDPKPYEQOGNVID 752
 Qy 602 FYGA-----EGEIFY-----KETPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFI 649
 Db 753 FDSVPQIQGQNNQNSQNSFEEDTEEDKPKYEQG--GNIIDIDFSDVPQIHG-----FNKHEI 806
 Qy 650 AQDONAPRVPAAKLGPH 667
 Db 807 IEDTNKDK-PNYQFGGH 823

RESULT 34

US-09-815-242-12686
 ; Sequence 12686, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.

```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12686
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12686

Query Match      2.9%; Score 117; DB 9; Length 1001;
Best Local Similarity 18.9%; Pred. NO. 0.14;
Matches 162; Conservative 103; Mismatches 277; Indels 316; Gaps 44;

Qy 19 PLLAQAHETQSGVLETV---TVGKSRBRATSGLLHTSTASDKLIISGDTLRQKAVNLGD 75
Db 96 FKAVQAPQTAQPLANLEKEVVKEAKFKQ---VKETIQPDN---SGD---QKQVDL-- 144
Qy 76 ALDQVGFQIHASQYGGASAPVIRGQTRRIKVLNHHGETGMADFPSPHAINMVDLTALSCQ 135
Db 145 -----TPKMATQ-----NQAAET-----Q 158
Qy 136 VEILRGFVTLVSSGNVAGLVADGK-IPEKMPENGVSGLGLRLSSGNLEKLTSGGIN 194
Db 159 VEAQ-Q-PRVTSNPRVTSADVVVEAKEASDAKVEGTGTVTSKVTVESGSIAPQGNKVE 217
Qy 195 IGLGRNPLV-----HTEGLYRKSGD-----YAVPRYRLKRLPDSPPRFANGQ 237
Db 218 PHAQORVLKYKLFKEKGLHK--GDYFDFTLSSNNVTVGYSTARKVPEIKNGSVVMATQ 275
Qy 238 HRAVLGWRKFRYRTYSDRDQYGLPAHSHEDDCHADLIWQKSLINKRYLQLYPHLATE 297
Db 276 ---LLGNGK--IATFTDYID-YKVNVT-----DLBINIFIDPKTVQSGNQQTITSTLND 325
Qy 298 ED-----VDYDPLGSCGFHDDDDAHAAHNGKPMIDLNRKYE----- 336
Db 326 KETKNTLPIEY-NFGVS-----NSYANVNGSIETFDKGNRFTHVAYIKPQNGHKSDS 377
Qy 337 -----LRAEWKQP-FPGFEALRVHLNRNDYHDEKAGDAVENFENNOTQARTEL 385
Db 378 VSTIGTLTGSKADGKAPTQKVEVLK---DANELFQSVYANVSDSSMPFADVTEENKOKL 434
Qy 386 RHQ-----PIGRKGSWGVOYLQKSSALSATSBAVKQPMLLDNKVQHYFFGVE--- 435
Db 435 KVENNGNYKLDIEKLEKSYVIHVDGE---VLSGSDQVNFRTMFGYPEQYKYYTHLGY 491
Qy 436 QANVDN-----FLEGQVRVEKQASIRTDKALIDR-ENY 469
Db 492 KLTWDNGLVLYSNKAKGDGTNGTITENNNMTFDDEYGTGV-----ITGOYDKNLVTTVEE 547

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12686
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12686

Query Match      2.9%; Score 115.5; DB 9; Length 981;
Best Local Similarity 19.3%; Pred. No. 0.19;
Matches 141; Conservative 129; Mismatches 265; Indels 197; Gaps 38;

Qy 86 SQVGGASAPVIRGQTRRIKVLNHHGET-GNADSPHAINMVDLTALSCQVEILRGPT 144
Db 69 NOYGGK---VLIDDDGLEIEVERIKGSAQGDVKVYLPNGAVRDDAWLQKLNYSKKT 125

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12211
; LENGTH: 981
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12211

Query Match      2.9%; Score 115.5; DB 9; Length 981;
Best Local Similarity 19.3%; Pred. No. 0.19;
Matches 141; Conservative 129; Mismatches 265; Indels 197; Gaps 38;

Qy 86 SQVGGASAPVIRGQTRRIKVLNHHGET-GNADSPHAINMVDLTALSCQVEILRGPT 144
Db 69 NOYGGK---VLIDDDGLEIEVERIKGSAQGDVKVYLPNGAVRDDAWLQKLNYSKKT 125

```

QY 145 LLXSSGNVAGLVADGKIPEKMPENGVSGBGLRLSSGNLEKLTSGINIGLGKQNFV 202
 Db 126 QGIFSDVLGLQDI-HRNLNEKQLQYLLQAGALGSTFTSMREVNR 172
 QY 203 LHTEGLYKSG--DYAVPRYRNK--RLPD--SPRFANGQHR 239
 Db 173 -KDELYKSGKMPIINQOIEQLKQLESQIREEAKLETYHRLVDORXSRLEMLKGN 231
 QY 240 AVLGRKFRYRTYSDRRDQYGLPAKSHEDDCHADIWOKSLINKRYLQYLPHLLTEED 299
 Db 232 L-----NOLSKMHEEKQKVALHDSQEW-----KSL--BQQLNIPITPEKG 273
 QY 300 VD-YDNFGLSCGFHDDDAHAHNGKPMIDLRKRY-ELRAEMKQFPF----- 346
 Db 274 VDRYEK-----ARAHKQSLERDGLRNERLAQKEATOLEPVKQSDIDAFIS 321
 QY 347 -----GFE-----ALRVHLNRNDYHDEKAGDAVENFFNNQ7Q 379
 Db 322 LNOQENEIKKEFELTAIEDKIANKQKQKDELQANIGWSETHDVSSEAMKSYVSEQIK 381
 QY 380 NARIELRHQPIGLKSGWQYLGQKSSALSATSEAVKQPMLLD---NKVQVYSPFVEQ 436
 Db 382 NKQOAA--IKQLERSLEENKI--EDNAVHSELDSVEEKIVPEETFEKKKEYSQQVIEL 437
 QY 437 ANWDF--TLEGGVVRKOKASIRYDICALIDRENYKQPLDILGAHROTARSPALSQWY 494
 Db 438 NEKENLYSKLKERFEIEQEKOKR-QKLL--RTTFILLTLVGIGL--TAFSP-IGNNML 490
 QY 495 F-----TPQKLSLTASHQERLPSTOELYANGKHVATNTFVGNGH---LNKERSNNI 544
 Db 491 FGIIFAVLTVFVIGIMSKSEVYSEAL-----TDEIBEIKQALIDENYDLDF 542
 QY 545 ELALGYE-GDRWQYNL-----ALYRNFRGNVYIAQTLDGRGPKSIEDSEMKLVRYNQSG 599
 Db 543 DLDEQYRDRHWQALKNKQILEKR-QYTEGR-LNDKAG-----RHDELQ 586
 QY 600 ADFVCAEGEIEYKTPRYRIGVSGDYVRGLKNLPSLPGRE-----DAYGNRPFIADQ 652
 Db 587 STVENKDELKLSK-----ISNDLIVDSISTWANTKALDQIHSDINQORQLVQELDT 640
 QY 653 -DONAPRVFAARLGFHLKASLTDR-----DANLDYRVFAQNKLARYETRTFQHHM 703
 Db 641 FYNHAEAVTKSQVYFNKLSLFDVQWKLKSAEDTN-EKWRINAEN-----IKLVNTEL 693
 QY 704 LNLGNRYRNTR 715
 Db 694 NHLNAQLENNK 705

RESULT 36
 US-10-260-877-62
 ; Sequence 62, Application US/10260877
 ; Publication No. US20030021813A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Abbott Laboratories
 ; APPLICANT: Chovan, Linda E.
 ; APPLICANT: Hessler, Paul E.
 ; APPLICANT: Reich, Karl A.
 ; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
 ; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
 ; TITLE OF INVENTION: 'ESSENTIAL GENES'
 ; FILE REFERENCE: 6565.US.P1
 ; CURRENT APPLICATION NUMBER: US/10/260.877
 ; CURRENT FILING DATE: 2002-09-30
 ; PRIOR APPLICATION NUMBER: US/09/649.145
 ; PRIOR FILING DATE: 2000-08-25
 ; NUMBER OF SEQ ID NOS: 137
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 62
 ; LENGTH: 631
 ; TYPE: PRF
 ; ORGANISM: H. influenzae
 ; US-10-260-877-62

Query Match 2.8%; Score 112; DB 15; Length 631;
 Best Local Similarity 18.7%; Pred. No. 0.2;
 Matches 128; Conservative 76; Mismatches 240; Indels 240; Gaps 35;
 QY 154 GLVDVADGKIPEKMPENGVSGBGLRLSSGNLEKLTSGINIGLGKQNF 201
 Db 80 GTITISONGI--GWTREQVIDHLGTIAKSGTEFLTALGQDQAKNSQLIQFGVGYSAF 137
 QY 202 VLHTEGLYRK-----SGDYAVPRYRNKRLPDSPPRFANGQHRVAVLGM 244
 Db 138 IVADKVTVKTRAAGEADKAVLWESAGEGEYSVADIEKKRGTDVILHLREDEKEFLNEW 197
 QY 245 RKRFRYRTYSDRRDQYGLPAH--SHEYDD---C-----HADIIWOKS---LINKRY 287
 Db 198 LRRIIIGKYS---DHIGLPVEMLTKEYDDGCKGCKWEKINKSDALWTRKNDVDEEY 254
 QY 288 LQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHNN-----GKPMIDLR 331
 Db 255 KAFYKHL-----SHDFVDPVTVAHNNKVEGNQAYTSLLYVPKAPW-DLF 297
 QY 332 NKRYELRAEWKQFPFPEALRVHLNRNDYHDEKAGDAVENFFNNQ7Q 389
 Db 298 NREHK-----HGLKLYVQVRFIMDD-----AEQFIPNLYRPMRGLSDNDLP 339
 QY 390 IGRLLKSGWQYLGQKSSALSATSEAVKQPMLLDNKVQHYSPFVEQANWDFNFTLEGVR 449
 Db 340 LN-----VSREILQDNKI-----TAALR 357
 QY 450 VEKOKASIRYDKALI--DRENY---YKQ-----PLPDLAGHROTARSPALSQWYFT 496
 Db 358 KALTKRSLOMLKLAQDAEKYLOFWKEFGLVLKGEPAEDF-ANKETVAKL----- 407
 QY 497 PQHKLSTASHQERLPSTQEL-----YAHGK-----HVAINTPEV--GNKHLNERSNNI 544
 Db 408 ----LRFASTINDGSEQTVSLEDIYLRKSGQKAIYIITADSYAAKNSPHLEFNKKG 463
 QY 545 E-LALGYEGDRWQYNLALYRNFRGNVYIAQTLDGRGPKSIEDSEMKLVRYNQSGADFY 603
 Db 464 EYLLLSDRIDEMWLS---YLTEF-----DGKQLOSI-TKADLDL-----GDLA 502
 QY 604 GAEGEIEYKTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFIADQDQNAAPRVFAAR 663
 Db 503 DKSE-----TQKQDEAFSGFIE-RVKNL-----LGERVKTVRLTHNLTDP-- 545
 QY 664 LGFHLKASLTDRIDANLDYRVFAQNKLARYETR-----TPGHMLNLGNRYRNTRYGE 718
 Db 546 -----VVSTDNQWTTQMAKLFRAAGQVPVEVYTFELNPEHHLVKKVADIADETE 599
 QY 719 WNVYVYKADNLLNQSVYAHSSFLSD 742
 Db 600 W-----VELLLEQAMLAERGSLEN 618

RESULT 37
 US-09-738-626-6493
 ; Sequence 6493, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738.626
 ; CURRENT FILING DATE: 2000-12-18


```

; TITLE OF INVENTION: Gamma Herpes Virus DNA and Methods of Use
; FILE REFERENCE: 61750-379
; CURRENT APPLICATION NUMBER: US/10/055,364
; CURRENT FILING DATE: 2002-01-23
; PRIOR APPLICATION NUMBER: US/09/612,204
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US/60/142,736
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: US/60/168,532
; PRIOR FILING DATE: 1999-12-02
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 618
; TYPE: PR1
; ORGANISM: Equine herpesvirus 5
US-10-055-364-44

```

[illegible]

RESULT 39
US-10-022-461-5
; Sequence 5, Application US/10022461
; Publication No. US20020142420A1
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.

Fri Nov 21 10:34:26 2003

us-09-936-377-2.rapb

Page 27

Search completed: November 14, 2003, 11:05:46
Job time : 47 secs
